

WIP

 WIP (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular

MSPrch_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:41:02 2000; Maspar time 3.04 Seconds

Tabular output not generated. 116.737 Million cell updates/sec

Title: >US-08-991-628-2
 Description: (1-15) from US08991628.pep
 Perfect Score: 105
 Sequence: 1 FGIFVVDKNTGDIINI 15

Scoring table: PAM 150
 Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries
 Maximum DB seq length 15

Database: a:geneseg35
 1:genesegp

Statistics: Mean 18.538; Variance 60.235; scale 0.308

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description | Pred. No. |
|------------|-------|-------------|-----------|----------|---------------------------------|-----------|
| 1 | 105 | 100.0 | 15 | 1 W04842 | Self epitope of desmog 3.51e-04 | |

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
 ID W04842 standard: peptide: 15 AA.
 AC W04842:
 DE Self epitope of desmoglein 3, implicated in autoimmune disease.
 DT 18-FEB-1997 (first entry)
 KW autoantigen; self-epitope; antigen; autoimmune disease;
 KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
 KW pemphigus vulgaris; desmoglein; multiple sclerosis;
 KW herpes simplex virus; adenovirus; phosphomannomutase;
 KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
 KW Influenza; haemagglutinin; reovirus; sigma protein.
 OS Homo sapiens.
 PN W09627387-A1.
 PD 12-SEP-1996.
 PF 07-MAR-1996; U03182.

PR 07-MAR-1995; US-400796.
 PA (HARD) HARVARD COLLEGE.
 PI Strominger JL, Muecherfennig KW;
 DR WPI: 96-425218/42.
 PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
 PT antigens - useful in disease treatment, and method for
 PT identification of other self and non-self antigens implicated in
 PT auto-immune disease
 PS Claim 1: Page 38; 58pp; English.
 CC Pharmaceutical preparations for tolerisation to antigens comprise
 CC either an isolated human non-collagen or non-mysin basic protein
 CC (MBP) polypeptide which is capable of tolerising an individual to an
 CC autoantigen; or an isolated human pathogen polypeptide capable of
 CC tolerising an individual to that polypeptide. In both cases, the
 CC polypeptide (whether self or non-self) includes an amino acid
 CC sequence corresponding to a sequence motif for a MHC class II
 CC protein, such as HLA-DR, which is associated with a human autoimmune
 CC disease and which binds to the polypeptide to activate autoreactive
 CC T-cells in individuals with the autoimmune disease. This peptide is
 CC derived from the human desmoglein 3 protein (amino acids 97-111)
 CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
 CC derived from the human desmoglein protein are described in W04841-47.
 SQ Sequence 15 AA:

Query Match 100.0%; Score 105; DB 1; Length 15;
 Best local Similarity 100.0%; Pred. No. 3.51e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 FGIFVVDKNTGDIINI 15
 ||||||||||||
 QY 1 FGIFVVDKNTGDIINI 15

Search completed: Sat May 13 04:41:09 2000
 Job time : 7 secs.

MIPS
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch.p protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:15:33 2000; Maspar time 401.42 Seconds

Tabular output not generated. 4.171 Million cell updates/sec

Title: >US-08-991-628-2

Description: (1-15) from US08991628.pep

Sequence: 1 FGIFVVDKNTGDINI 15

Scoring table: PAM 150
Gap 15

Searched: 714217 seqs, 111628056 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries
Maximum DB seq length 15

Database: a-pending

1:PCT 2:06 3:060 4:07 5:080 6:081 7:082 8:083 9:084A
10:084B 11:085 12:086 13:087 14:088 15:089 16:090 17:091
18:092 19:093 20:094 21:095 22:NEWP 23:NEWU60 24:NEWU8
25:NEWU9

Statistics: Mean 21.138; Variance 58.541; scale 0.361

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|-------------|------------------------|
| 1 | 105 | 100.0 | 15 | 1 | PCT-US96-0 | Sequence 2, Applicatio |
| 2 | 105 | 100.0 | 15 | 15 | US-08-991- | Sequence 2, Applicatio |

Note: Post-processor removed 43 summaries from list due to search parameters chosen.

ALIGNMENTS

| RESULT | ID | PCT-US96-03182-2 | STANDARD: | PRT: | 15 AA. |
|--------|----|---------------------------------------|-----------|------|--------|
| XX | AC | xxxxxx | | | |
| XX | DE | Sequence 2, Application PC/TUS9603182 | | | |
| XX | CC | Sequence 2, Application PC/TUS9603182 | | | |
| CC | CC | GENERAL INFORMATION: | | | |

| | | |
|--------|----|---|
| CC | CC | APPLICANT: |
| CC | CC | TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF |
| CC | CC | TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE |
| CC | CC | NUMBER OF SEQUENCES: 16 |
| CC | CC | CORRESPONDENCE ADDRESS: |
| CC | CC | ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C. |
| CC | CC | STREET: 600 ATLANTIC AVENUE |
| CC | CC | CITY: BOSTON |
| CC | CC | STATE: MASSACHUSETTS |
| CC | CC | COUNTRY: UNITED STATES OF AMERICA |
| CC | CC | ZIP: 02210 |
| CC | CC | COMPUTER READABLE FORM: |
| CC | CC | MEDIUM TYPE: Floppy disk |
| CC | CC | COMPUTER: IBM PC compatible |
| CC | CC | OPERATING SYSTEM: PC-DOS/MS-DOS |
| CC | CC | SOFTWARE: PatentIn Release #1.0, Version #1.25 |
| CC | CC | CURRENT APPLICATION DATA: |
| CC | CC | APPLICATION NUMBER: PCT/US96/03182 |
| CC | CC | FILING DATE: |
| CC | CC | CLASSIFICATION: |
| CC | CC | PRIOR APPLICATION DATA: |
| CC | CC | APPLICATION NUMBER: US 08/400,796 |
| CC | CC | FILING DATE: 07-MAR-1995 |
| CC | CC | ATTORNEY/AGENT INFORMATION: |
| CC | CC | NAME: GATES, EDWARD R. |
| CC | CC | REGISTRATION NUMBER: 31,616 |
| CC | CC | REFERENCE/DOCKET NUMBER: H0498/7015MO |
| CC | CC | TELECOMMUNICATION INFORMATION: |
| CC | CC | TELEPHONE: 617-720-3500 |
| CC | CC | TELEFAX: 617-720-2441 |
| CC | CC | INFORMATION FOR SEQ ID NO: 2: |
| CC | CC | SEQUENCE CHARACTERISTICS: |
| CC | CC | LENGTH: 15 amino acids |
| CC | CC | TYPE: amino acid |
| CC | CC | TOPOLOGY: linear |
| CC | CC | MOLECULE TYPE: peptide |
| CC | CC | HYPOTHETICAL: NO |
| CC | CC | FRAGMENT TYPE: Internal |
| CC | CC | ORIGINAL SOURCE: |
| CC | CC | ORGANISM: HOMO SAPIENS |
| CC | CC | SEQUENCE 15 AA; 1652 MW; 1115 CN; |
| DB | DB | 1 FGIFVVDKNTGDINI 15 |
| OY | OY | 1 FGIFVVDKNTGDINI 15 |
| RESULT | 2 | |
| ID | ID | US-08-991-628-2 |
| XX | XX | xxxxxx |
| XX | XX | Sequence 2, Application US/08991628 |
| XX | XX | Sequence 2, Application US/08991628 |
| CC | CC | GENERAL INFORMATION: |
| CC | CC | APPLICANT: STROMINGER, Jack L |
| CC | CC | APPLICANT: WICHERPENNING, Kai |
| CC | CC | TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF ANTIGENS IMPLICATED |
| CC | CC | TITLE OF INVENTION: IN AUTOIMMUNE DISEASE |
| CC | CC | FILE REFERENCE: HAR-001DV |
| CC | CC | CURRENT APPLICATION NUMBER: US/08/991,628 |
| CC | CC | CURRENT FILING DATE: 1997-11-15 |
| CC | CC | EARLIER APPLICATION NUMBER: USSN 08/400,796 |
| CC | CC | EARLIER FILING DATE: 1995-03-07 |
| CC | CC | NUMBER OF SEQ ID NOS: 20 |
| CC | CC | SOFTWARE: PatentIn Ver. 2.0 |

CC SEQ ID NO 2
 CC LENGTH: 15
 CC TYPE: PRT
 CC ORGANISM: Homo sapiens
 SQ SEQUENCE 15 AA: 1652 MW: 1115 CN:

Query Match 100.0%; Score 105; DB 15; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.57e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 FGIFVVDKNTGDINI 15
 QY 1 FGIFVVDKNTGDINI 15

Search completed: Sat May 13 05:22:31 2000
 Job time : 418 secs.

WOLFE (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_DP protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:14:06 2000; MasPar time 55.78 Seconds

Tabular output not generated. 3.484 Million cell updates/sec

Title: >US-08-991-628-2

Description: (1-15) from US08991628.pap

Perfect Score: 105

Sequence: 1 FGIFVVDKNTGDIINI 15

Scoring table: PAM 150

Gap 15

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Maximum DB seq length 15

Database: a-issued

1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 17.421; Variance 58.153; scale 0.300

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description | Pred. No. |
|------------|-------|-------------|-----------|----|-----------------------------------|-----------|
| 1 | 105 | 100.0 | 15 | 2 | US-08-400- Sequence 2, Applicatio | 2.01e-04 |

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID US-08-400-796-2 STANDARD; PRT: 15 AA.
XX
AC xxxxxx
XX
DT
XX
XX Sequence 2, Application US/08400796
XX
CC Sequence 2, Application US/08400796
CC Patent No. 5874531
CC GENERAL INFORMATION:
CC APPLICANT: STROMINGER, JACK L.
CC APPLICANT: WUCHERPFENNIG, KAI
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF

CC TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC STREET: 600 ATLANTIC AVENUE
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/400,796
CC FILING DATE:
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: TWOMEY, MICHAEL J.
CC REGISTRATION NUMBER: 38,349
CC REFERENCE/DOCKET NUMBER: H0498/7015
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-3500
CC TELEFAX: 617-720-2441
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: internal
CC ORIGINAL SOURCE:
CC ORGANISM: HOMO SAPIENS
CC SEQUENCE 15 AA: 1652 MW: 1115 CN:
DB 1 FGIFVVDKNTGDIINI 15
Oy 1 FGIFVVDKNTGDIINI 15
Query Match 100.0%; Score 105; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.01e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: Sat May 13 05:15:11 2000
Job time : 65 secs.

WISN (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:40:34 2000; Maspar time 4.34 Seconds
162.858 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-2
Description: (1-15) from US08991628.pep
Perfect score: 105
Sequence: 1 FGIFVVDKNTGDINI 15

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 26.968; Variance 38.706; scale 0.697.

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Length | ID | Description | Pred. No. |
|------------|-------------|--------|-------|-------------|-----------|
| | | | | | |

No matches found.

Search completed: Sat May 13 04:40:42 2000
Job time : 8 secs.

 WIDEORL (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:34:04 2000; Maspar time 101.68 Seconds

4.493 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-2
 Description: (1-15) from US08991628.pep
 Perfect Score: 105
 Sequence: 1 FGIFVVDKMTGDINI 15

Scoring table:
 PAM 150
 Gap 15

Searched: 83857 segs, 30454973 residues

Post-processing: Minimum Match 08
 Listing first 45 summaries
 Maximum DB seq length 15

Database: swiss-prot38
 1:swissprot

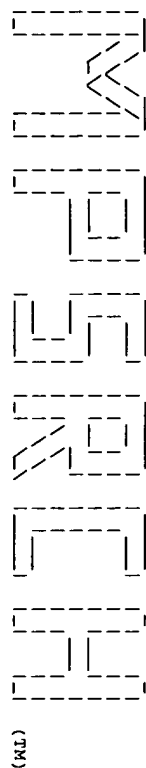
Statistics: Mean 27.553; Variance 34.618; scale 0.796

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Score | Match Length | ID | Description | Pred. No. |
|-------------------|-------------|--------------|----|-------------|-----------|
| No matches found. | | | | | |

Search completed: Sat May 13 04:35:56 2000
 Job time : 112 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:36:17 2000: Mpsrch time 223.46 seconds

Tabular output not generated. 4.654 Million cell updates/sec

Title: >US-08-991-628-2
Description: (1-15) from US08991628.pep
Perfect Score: 105
Sequence: 1 FGIFFVVDKNTGDIINI 15

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 26.732; Variance 34.969; scale 0.764

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.







SUMMARIES

| Result No. | Query Match | Score | Length | ID | Description | Pred. No. |
|------------|-------------|-------|--------|----|-------------|-----------|
| | | | | | | |

No matches found.

Search completed: Sat May 13 04:40:15 2000
Job time : 238 secs.

THIS PAGE BLANK (uspto)

M P E S E R E H
(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_p protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:22:51 2000: Maspar time 52.56 Seconds
3.698 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-3

Description: (1-15) from US08991628.pep

Perfect Score: 96

Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: PAM 150
Gap 15

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Maximum DB seq length 15

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfile1

Statistics: Mean 16.952; Variance 50.735; scale 0.334

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Score | Match Length | DB ID | Description | Pred. No. |
|------------|-------------|--------------|-------|-------------------------------------|-----------|
| 1 | 96 | 100.0 | 15 | 2 US-08-400- Sequence 3, Applicatio | 2.73e-04 |
| 2 | 96 | 100.0 | 15 | 1 US-08-787- Sequence 24, Applicati | 2.73e-04 |

Note: Post-processor removed 43 summaries from list due to search parameters chosen.

ALIGNMENTS

| RESULT ID | 1 | STANDARD: | PRT: | 15 AA. |
|-----------|-------------------------------------|-----------|------|--------|
| XX | US-08-400-796-3 | | | |
| AC | xxxxxx | | | |
| XX | | | | |
| DT | | | | |
| XX | | | | |
| DE | Sequence 3, Application US/08400796 | | | |
| XX | | | | |
| CC | Sequence 3, Application US/08400796 | | | |
| CC | Patent No. 5874531 | | | |
| CC | GENERAL INFORMATION: | | | |
| CC | APPLICANT: STROMINGER, JACK L. | | | |
| CC | APPLICANT: WOCHERPFENNIG, KAI | | | |

| | |
|----|---|
| CC | TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF |
| CC | TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE |
| CC | NUMBER OF SEQUENCES: 16 |
| CC | CORRESPONDENCE ADDRESS: |
| CC | ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C. |
| CC | STREET: 600 ATLANTIC AVENUE |
| CC | CITY: BOSTON |
| CC | STATE: MA |
| CC | COUNTRY: USA |
| CC | ZIP: 02210 |
| CC | COMPUTER READABLE FORM: |
| CC | MEDIUM TYPE: Floppy disk |
| CC | COMPUTER: IBM PC compatible |
| CC | OPERATING SYSTEM: PC-DOS/MS-DOS |
| CC | SOFTWARE: Patentin Release #1.0, Version #1.25 |
| CC | CURRENT APPLICATION DATA: |
| CC | APPLICATION NUMBER: US/08/400,796 |
| CC | FILING DATE: |
| CC | CLASSIFICATION: 424 |
| CC | ATTORNEY/AGENT INFORMATION: |
| CC | NAME: TROMER, MICHAEL J. |
| CC | REGISTRATION NUMBER: 38,349 |
| CC | REFERENCE/DOCKET NUMBER: H0498/7015 |
| CC | TELECOMMUNICATION INFORMATION: |
| CC | TELEPHONE: 617-720-3500 |
| CC | TELEFAX: 617-720-2441 |
| CC | INFORMATION FOR SEQ ID NO: 3: |
| CC | SEQUENCE CHARACTERISTICS: |
| CC | LENGTH: 15 amino acids |
| CC | TYPE: amino acid |
| CC | TOPOLOGY: linear |
| CC | MOLECULE TYPE: peptide |
| CC | HYPOTHETICAL: NO |
| CC | FRAGMENT TYPE: Internal |
| CC | ORIGINAL SOURCE: |
| CC | ORGANISM: HOMO SAPIENS |
| CC | SEQUENCE 15 AA: 1645 MW; 1217 CN; |
| DB | 1 LNSKIAFKIVSQEPA 15 |
| DB | 1 LNSKIAFKIVSQEPA 15 |
| Qy | 1 LNSKIAFKIVSQEPA 15 |
| DE | Sequence 24, Application US/08787547 |
| XX | |
| AC | xxxxxx |
| XX | |
| DT | |
| XX | |
| DE | Sequence 24, Application US/08787547 |
| XX | |
| CC | Patent No. 5783567 |
| CC | GENERAL INFORMATION: |
| CC | APPLICANT: Hedley, Mary Lynne |
| CC | APPLICANT: Curley, Joanne M. |
| CC | APPLICANT: Langer, Robert S. |
| CC | TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY |
| CC | TITLE OF INVENTION: OF NUCLEIC ACID |
| CC | NUMBER OF SEQUENCES: 107 |
| CC | CORRESPONDENCE ADDRESS: |
| CC | ADDRESSEE: Fish & Richardson, P.C. |
| CC | STREET: 225 Franklin Street |
| CC | CITY: Boston |
| CC | STATE: MA |
| CC | COUNTRY: US |
| CC | ZIP: 02110-2804 |
| CC | COMPUTER READABLE FORM: |

CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: Windows95
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/787,547
CC FILING DATE: 22-JAN-1997
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fraser, Janis K.
CC REGISTRATION NUMBER: 34,819
CC REFERENCE/DOCKET NUMBER: 08191/003001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-542-5070
CC TELEFAX: 617-542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 24:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 15 AA; 1645 MW; 1217 CN;

Query Match 100.00; Score 96; DB 1; Length 15;
Best Local Similarity 100.00; Pred. No. 2.73e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 LNSKIAFKIVSOEPA 15
|||
QY 1 LNSKIAFKIVSOEPA 15

Search completed: Sat May 13 05:23:57 2000
Job time : 66 secs.

M P S R C H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:24:16 2000: Maspar time 398.53 seconds

Tabular output not generated.

Title: >US-08-991-628-3

Description: (1-15) from US08991628.pep

Perfect Score: 96

Sequence: 1 LNSKIAFKIVSOEPA 15

Scoring table:

PAM 150

Searched: 714217 seqs, 111628056 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Maximum DB seq length 15

Database: a-pending

1:PCT 2:U6 3:U60 4:U7 5:U80 6:U81 7:U82 8:U83 9:U84A

10:U84B 11:U85 12:U86 13:U87 14:U88 15:U89 16:U90 17:U91

18:U92 19:U93 20:U94 21:U95 22:NEWP 23:NEWU60 24:NEWU8

25:NEWU9

Statistics: Mean 20.540; Variance 50.756; scale 0.405

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|-----------------------------------|-----------|
| 1 | 96 | 100.0 | 15 | 19 | US-09-321- Sequence 24, Applicat | 5.13e-04 |
| 2 | 96 | 100.0 | 15 | 16 | US-09-003- Sequence 24, Applicat | 5.13e-04 |
| 3 | 96 | 100.0 | 15 | 1 | PCT-US96-0 Sequence 3, Applicatio | 5.13e-04 |
| 4 | 96 | 100.0 | 15 | 15 | US-08-991- Sequence 3, Applicatio | 5.13e-04 |
| 5 | 96 | 100.0 | 15 | 1 | PCT-US98-0 Sequence 24, Applicati | 5.13e-04 |

Note: Post-processor removed 40 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID US-09-321-346-24 STANDARD: PRT: 15 AA.

AC xxxxxx

XX

DE Sequence 24, Application US/09321346

XX Sequence 24, Application US/09321346
CC GENERAL INFORMATION:
CC APPLICANT: Lunsford, Lynn B.
CC APPLICANT: Putnam, David
CC APPLICANT: Hedley, Mary Lynn
CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF NUCLEIC ACID
CC FILE REFERENCE: 08191/014001
CC CURRENT APPLICATION NUMBER: US/09/321,346
CC CURRENT FILING DATE: 1999-05-27
CC EARLIER APPLICATION NUMBER: US 09/266,463
CC EARLIER FILING DATE: 1999-03-11
CC NUMBER OF SEQ. ID NOS: 114
CC SOFTWARE: FastSeq for Windows Version 3.0
CC SEQ ID NO 24
CC LENGTH: 15
CC TYPE: PRT
CC ORGANISM: Homo sapiens
CC SEQUENCE 15 AA: 1645 MW: 1217 CN:
S0
Db 1 LNSKIAFKIVSOEPA 15
OY 1 LNSKIAFKIVSOEPA 15
Query Match 100.0%; Score 96; DB 19; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.13e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
RESULT 2
ID US-09-003-253-24 STANDARD: PRT: 15 AA.
AC xxxxxx
XX
DT
XX
DE Sequence 24, Application US/09003253
CC
CC GENERAL INFORMATION:
CC APPLICANT: Hedley, Mary Lynn
CC APPLICANT: Curley, Joanne M.
CC APPLICANT: Langer, Robert S.
CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF NUCLEIC ACID
CC NUMBER OF SEQUENCES: 108
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson P.C.
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: Windows 95
CC SOFTWARE: FastSeq for Windows Version 2.0b
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/003,253
CC FILING DATE: 06-JAN-1998
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/035,983
CC FILING DATE: 22-JAN-1997
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fraser, Ph.D., J.D., Janis K.
CC REGISTRATION NUMBER: 34,819
CC REFERENCE/DOCKET NUMBER: 08191/003002
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617/542-5070
CC TELEFAX: 617/542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 24:

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 15 AA: 1645 MW: 1217 CN:
SQ

Query Match 100.0%; Score 96; DB 16; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.13e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 LNSKIAFKIVSOEPA 15
OY 1 LNSKIAFKIVSOEPA 15

RESULT 3
ID PCT-US96-03182-3 STANDARD; PRT: 15 AA.
XX xxxxxx
XX

Sequence 3, Application PC/TUS9603182
CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF
CC TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC STREET: 600 ATLANTIC AVENUE
CC CITY: BOSTON
CC STATE: MASSACHUSETTS
CC COUNTRY: UNITED STATES OF AMERICA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC OPERATING SYSTEM: IBM PC COMPATIBLE
CC SOFTWARE: PC-DOS/MS-DOS
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/03182
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/400,796
CC FILING DATE: 07-MAR-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: GATES, EDWARD R.
CC REGISTRATION NUMBER: 31,616
CC REFERENCE/DOCKET NUMBER: H0498/7015WO
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-3500
CC TELEFAX: 617-720-2441
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: internal
CC ORIGINAL SOURCE:
CC ORGANISM: HOMO SAPIENS
CC SEQUENCE 15 AA: 1645 MW: 1217 CN:
SQ

Query Match 100.0%; Score 96; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.13e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 LNSKIAFKIVSOEPA 15
OY 1 LNSKIAFKIVSOEPA 15

RESULT 4
ID US-08-991-628-3 STANDARD; PRT: 15 AA.
XX xxxxxx
XX

Sequence 3, Application US/08991628
CC GENERAL INFORMATION:
CC APPLICANT: STROMINGER, Jack L
CC APPLICANT: WUCHERPPENNIG, Kai
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF ANTIGENS IMPLICATED
CC FILE REFERENCE: HAR-001DV
CC CURRENT APPLICATION NUMBER: US/08/991,628
CC CURRENT FILING DATE: 1997-11-15
CC EARLIER APPLICATION NUMBER: USSN 08/400,796
CC EARLIER FILING DATE: 1995-03-07
CC NUMBER OF SEQ ID NOS: 20
CC SOFTWARE: PatentIn Ver. 2.0
CC SEQ ID NO 3
CC LENGTH: 15
CC TYPE: PRT
CC ORGANISM: Homo sapiens
SQ SEQUENCE 15 AA: 1645 MW: 1217 CN:
SQ

Query Match 100.0%; Score 96; DB 15; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.13e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 LNSKIAFKIVSOEPA 15
OY 1 LNSKIAFKIVSOEPA 15

RESULT 5
ID PCT-US98-01499-24 STANDARD; PRT: 15 AA.
XX xxxxxx
XX

Sequence 24, Application PC/TUS9801499
CC GENERAL INFORMATION:
CC APPLICANT: Pangaea, Inc.
CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY
CC TITLE OF INVENTION: OF NUCLEIC ACID
CC NUMBER OF SEQUENCES: 107
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson, P.C.
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: US
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC OPERATING SYSTEM: Windows95
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US98/01499
CC FILING DATE: 22-JAN-1998
CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 08/787,547
 CC FILING DATE: 22-JAN-1997
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Fraser, Janis K.
 CC REGISTRATION NUMBER: 34,819
 CC REFERENCE/DOCKET NUMBER: 08191/003WO1
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 617-542-5070
 CC TELEFAX: 617-542-8906
 CC TELEX: 200154
 CC INFORMATION FOR SEQ ID NO: 24:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 15 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC SEQUENCE 15 AA: 1645 MW: 1217 CN:
 SO
 Query Match 100.0%; Score 96; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 5,13e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 LNSKIAFKIVSOEPA 15
 ||||||||||||
 Oy 1 LNSKIAFKIVSOEPA 15

Search completed: Sat May 13 05:31:12 2000
 Job time : 416 secs.

WIDEORLH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_gp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 04:47:53 2000; MasPar time 4.31 Seconds
Tabular output not generated. 164.283 Million cell updates/sec

Title: >US-08-991-628-3
Description: (1-15) from US08991628.pap
Perfect Score: 96
Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: plr62
1:plr1 2:plr2 3:plr3 4:plr4

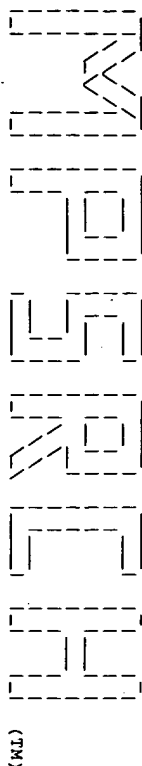
Statistics: Mean 25.750; Variance 32.305; scale 0.797

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query | No. Score Match | length DB | ID | Description | Pred. No. |
|-------------------|-------|-----------------|-----------|----|-------------|-----------|
| No matches found. | | | | | | |

Search completed: Sat May 13 04:48:01 2000
Job time : 8 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:41:28 2000; MasPar time 92.54 Seconds

Tabular output not generated. 4.936 Million cell updates/sec

Title: >US-08-991-628-3
Description: (1-15) from US08991628.pap
Perfect Score: 96
Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: PAM 150
Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: swiss-prot38
1:swissprot

Statistics: Mean 26.327; Variance 28.416; scale 0.926


Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES


| Result No. | Query Match | Length | ID | Description | Pred. No. |
|------------|-------------|--------|----|-------------|-----------|
| ***** | | | | | |

No matches found.

Search completed: Sat May 13 04:43:12 2000
Job time : 104 secs.







Release 3.1A John F. Collins, BioComputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on: Sat May 13 04:43:31 2000; MasPar time 228.30 Seconds
```

Tabular output not generated.

| | |
|----------------|----------------------------|
| Title: | >US-08-991-628-3 |
| Description: | (1-15) from US08991628.pep |
| Perfect Score: | 96 |
| Sequence: | 1 LNSKIAFKIVSQEPA 15 |

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

```
Post-processing:  Minimum Match 0%
                  Listing first 45 summaries
                  Maximum DB seq length 15
```

Database: **sprembl12**
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 25.549; Variance 30.093; scale 0.849

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query | ID | Description | Pred. No |
|--------|-------|--------------|-------------|----------|
| No. | Score | Match Length | DB | |
| | | | | |

No matches found.

Search completed: Sat May 13 04:47:34 2000
Job time : 243 secs.

THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSrch_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:33:05 2000; Maspar time 392.63 Seconds
4.265 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-4

Description: (1-15) from US08991628.pep

Sequence: 1 TPMFLSRNTGEVRT 15

Scoring table: PAM 150
Gap 15

Searched: 714217 segs, 111628056 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database:

a-pending
1:PCT 2:U6 3:U60 4:U7 5:U80 6:U81 7:U82 8:U83 9:U84A
10:U84B 11:U85 12:U86 13:U87 14:U88 15:U89 16:U90 17:U91
18:U92 19:U93 20:U94 21:U95 22:NEWP 23:NEW60 24:NEW08
25:NEW09

Statistics: Mean 21.981; Variance 55.748; scale 0.394

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|-----------------------------------|-----------|
| 1 | 108 | 100.0 | 15 | 15 | US-08-991- Sequence 4, Applicatio | 4.31e-05 |
| 2 | 108 | 100.0 | 15 | 1 | PCT-US96-0 Sequence 4, Applicatio | 4.31e-05 |
| 3 | 108 | 100.0 | 15 | 1 | PCT-US98-0 Sequence 25, Applicati | 4.31e-05 |
| 4 | 108 | 100.0 | 15 | 19 | US-09-321- Sequence 25, Applicati | 4.31e-05 |
| 5 | 108 | 100.0 | 15 | 16 | US-09-003- Sequence 25, Applicati | 4.31e-05 |
| 6 | 59 | 54.6 | 11 | 17 | US-09-187- Sequence 3275, Applic | 1.23e-02 |

Note: Post-processor removed 39 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID US-08-991-628-4 STANDARD: PRT: 15 AA.
AC xxxxxx
XX
DT
XX

DE Sequence 4, Application US/08991628
XX
CC Sequence 4, Application US/08991628
CC GENERAL INFORMATION:
CC APPLICANT: STROMINGER, Jack L
CC APPLICANT: WOCHERPFENNIG, Kai
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF ANTIGENS IMPLICATED
CC TITLE OF INVENTION: IN AUTOIMMUNE DISEASE
CC FILE REFERENCE: HAR-001DV
CC CURRENT APPLICATION NUMBER: US/08/991,628
CC CURRENT FILING DATE: 1997-11-15
CC EARLIER APPLICATION NUMBER: USSN 08/400,796
CC EARLIER FILING DATE: 1995-03-07
CC NUMBER OF SEQ ID NOS: 20
CC SOFTWARE: PatentIn Ver. 2.0
CC SEQ ID NO 4
CC LENGTH: 15
CC TYPE: PRT
CC ORGANISM: Homo sapiens
SQ SEQUENCE 15 AA; 1722 MW; 1303 CN;

Query Match 100.0%; Score 108; DB 15; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.31e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TPMFLSRNTGEVRT 15
OY 1 TPMFLSRNTGEVRT 15

RESULT 2
ID PCT-US96-03182-4 STANDARD: PRT: 15 AA.
XX
AC xxxxxx
XX
DT
XX

Sequence 4, Application PC/TUS9603182

Sequence 4, Application PC/TUS9603182
GENERAL INFORMATION:

APPLICANT:
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF
CC TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC STREET: 600 ATLANTIC AVENUE
CC CITY: BOSTON
CC STATE: MASSACHUSETTS
CC COUNTRY: UNITED STATES OF AMERICA
CC ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/03182
CC FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA: US 08/400,796
CC APPLICATION NUMBER: US 08/400,796
CC FILING DATE: 07-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.
CC REGISTRATION NUMBER: 31,616
CC REFERENCE/DOCKET NUMBER: HQ498/7015WO
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-3500
CC TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:


```

CC      LENGTH: 15 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: peptide
CC      HYPOTHETICAL: NO
CC      FRAGMENT TYPE: Internal
CC      ORIGINAL SOURCE:
CC      ORGANISM: HOMO SAPIENS
S0      SEQUENCE 15 AA: 1722 MW: 1303 CN:

Query Match      100.0%: Score 108; DB 1; Length 15;
Best Local Similarity 100.0%: Pred. No. 4.31e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 TPMFLSRNTGEVRT 15
        |||
0y      1 TPMFLSRNTGEVRT 15

RESULT 3
ID      PCT-US98-01499-25 STANDARD; PRT: 15 AA.
AC      xxxxxx
XX
XX
XX
XX
XX
XX      Sequence 25, Application PC/TUS9801499
DE
DE      Sequence 25, Application PC/TUS9801499
XX      GENERAL INFORMATION:
CC      APPLICANT: Pangaea, Inc.
CC      TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY
CC      TITLE OF INVENTION: OF NUCLEIC ACID
CC      NUMBER OF SEQUENCES: 107
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Fish & Richardson, P.C.
CC      STREET: 225 Franklin Street
CC      CITY: Boston
CC      STATE: MA
CC      COUNTRY: US
CC      ZIP: 02110-2804
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Diskette
CC      COMPUTER: IBM Compatible
CC      OPERATING SYSTEM: Windows95
CC      SOFTWARE: FASTSEQ for Windows Version 2.0
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: PCT/US98/01499
CC      FILING DATE: 22-JAN-1998
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: 08/787,547
CC      FILING DATE: 22-JAN-1997
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Fraser, Janis K.
CC      REGISTRATION NUMBER: 34,819
CC      REFERENCE/DOCKET NUMBER: 08191/003MO1
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: 617-542-5070
CC      TELEFAX: 617-542-8906
CC      TELEX: 200154
CC      INFORMATION FOR SEQ ID NO: 25:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 15 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: peptide
CC      SEQUENCE 15 AA: 1722 MW: 1303 CN:

Query Match      100.0%: Score 108; DB 1; Length 15;
Best Local Similarity 100.0%: Pred. No. 4.31e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 TPMFLSRNTGEVRT 15

```

```

0Y      1 TPFMFLSRNTEVRT 15
        |||
        |
RESULT  4
ID US-09-321-346-25 STANDARD: PRT: 15 AA.
XX xxxxxx
XX
XX
DE Sequence 25, Application US/09321346
CC
CC Sequence 25, Application US/09321346
CC GENERAL INFORMATION:
CC APPLICANT: Lunsford, Lynn B.
CC APPLICANT: Putnam, David
CC APPLICANT: Hedley, Mary Lynn
CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF NUCLEIC ACID
CC FILE REFERENCE: 08191/014001
CC CURRENT APPLICATION NUMBER: US/09/321,346
CC CURRENT FILING DATE: 1999-05-27
CC EARLIER APPLICATION NUMBER: US 09/266,463
CC EARLIER FILING DATE: 1999-03-11
CC NUMBER OF SEQ ID NOS: 14
CC SOFTWARE: FASTSEQ for Windows Version 3.0
CC SEQ ID NO 25
CC LENGTH: 15
CC TYPE: PRT
CC ORGANISM: Homo sapiens
SQ SEQUENCE 15 AA; 1722 MW; 1303 CN;

Query Match 100.0%; Score 108; DB 19; Length 15;
Best Local Similarity 100.0%; Pred. No. 4,31e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps

Db 1 TPFMFLSRNTEVRT 15
   |||
   |
OY 1 TPFMFLSRNTEVRT 15

RESULT  5
J US-09-003-253-25 STANDARD: PRT: 15 AA.
XX xxxxxx
XX
XX
DT
DE Sequence 25, Application US/09003253
XX
XX
CC Sequence 25, Application US/09003253
CC GENERAL INFORMATION:
CC APPLICANT: Hedley, Mary Lynne
CC APPLICANT: Curley, Joanne M.
CC APPLICANT: Langert, Robert S.
CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF NUCLEIC ACID
CC NUMBER OF SEQUENCES: 108
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Fish & Richardson P.C.
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: Windows 95
CC SOFTWARE: FASTSEQ for Windows Version 2.0b
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/003,253
CC FILING DATE: 06-JAN-1998
CC CLASSIFICATION: 514
```

CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 60/035,983
 CC FILING DATE: 22-JAN-1997
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Fraser, Ph.D., J.D., Janis K.
 CC REGISTRATION NUMBER: 34,819
 CC REFERENCE/DOCKET NUMBER: 08191/003002
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 617/542-5070
 CC TELEFAX: 617/542-8906
 CC TELEX: 200154
 CC INFORMATION FOR SEQ. ID NO: 25:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 15 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 SO SEQUENCE 15 AA; 1722 MW; 1303 CN;

Query Match 100.0%; Score 108; DB 16; Length 15;
 Best Local Similarity 100.0%; Pred. No. 4,31e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TPMFLSRNTGEVRT 15
 QY 1 TPMFLSRNTGEVRT 15

RESULT 6
 ID US-09-187-859-3275 STANDARD; PRT; 11 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX
 DE

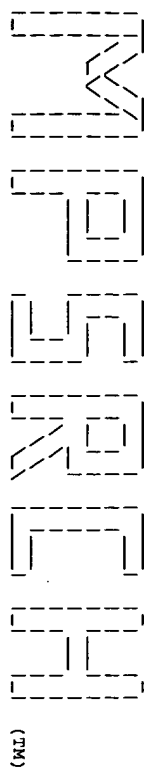
Sequence 3275, Application US/09187859

CC Sequence 3275, Application US/09187859
 CC GENERAL INFORMATION:
 CC APPLICANT: Blaschuk, Orest W.
 CC APPLICANT: Gour, Barbara J.
 CC TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
 CC TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
 CC FILE REFERENCE: 100086, 407C1
 CC CURRENT APPLICATION NUMBER: US/09/187,859
 CC CURRENT FILING DATE: 1998-11-06
 CC NUMBER OF SEQ ID NOS: 4052
 CC SOFTWARE: PatentIn Ver. 2.0
 CC SEQ ID NO 3275
 CC LENGTH: 11
 CC TYPE: PRT
 CC ORGANISM: Artificial Sequence
 CC FEATURE:
 CC OTHER INFORMATION: Representative cyclic modulating agent based on
 CC OTHER INFORMATION: Desmoglein cell adhesion recognition sequence
 SO SEQUENCE 11 AA; 1323 MW; 536 CN;

Query Match 54.6%; Score 59; DB 17; Length 11;
 Best Local Similarity 70.0%; Pred. No. 1.23e+02;
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 2 MFLINRNTGE 11
 QY 3 MFLSRNTGE 12

Search completed: Sat May 13 05:39:53 2000
 Job time : 408 secs.



 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:53:44 2000; MasPar time 4.32 Seconds 163.716 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-4
 Description: (1-15) from US08991628.pep
 Perfect Score: 108
 Sequence: 1 TPMFLSRNTEVRT 15

Scoring table: PAM 150
 Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 08
 Listing first 45 summaries
 Maximum DB seq length 15

Database: plr62
 1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 27.219; Variance 35.537; scale 0.766

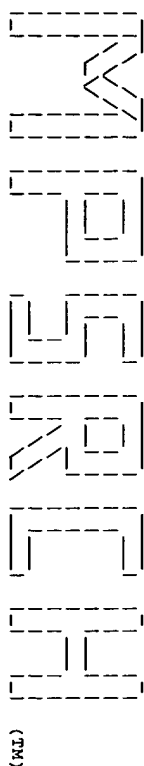
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | ID | Description | Pred. No. |
|------------|-------------|----|-------------|-----------|
| ----- | | | | |

No matches found.

Search completed: Sat May 13 04:53:53 2000
 Job time : 9 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:48:45 2000; MasPar time 90.27 Seconds
5.061 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-4
Description: (1-15) from US08991628.pep
Perfect Score: 108
Sequence: 1 TPMFLSRNTEGVRT 15

Scoring table: PAM 150
Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: swiss-prot38
1:swissprot

Statistics: Mean 27.727; Variance 32.559; scale 0.852

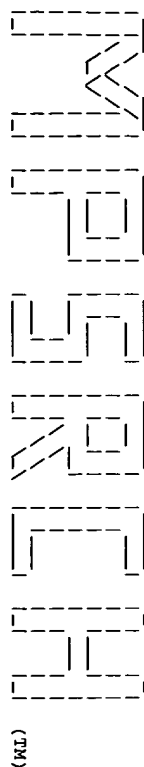
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query | | | | |
|--------|-------|-------|--------|----|-------------|
| No. | Score | Match | Length | ID | Description |
| | | | | | Pred. No. |

No matches found.

Search completed: Sat May 13 04:50:23 2000
Job time : 98 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPearch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:50:42 2000; Maspar time 149.09 Seconds
6.976 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-4
Description: (1-15) from US08991628.pep
Perfect Score: 108
Sequence: 1 TPFLLSRNTGEVRT 15

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_pnase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 27.384; Variance 31.843; scale 0.860

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

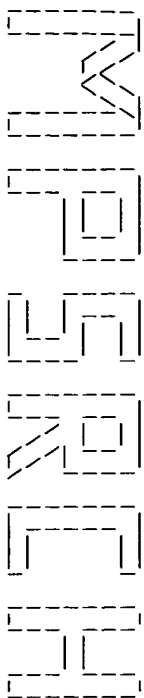
| Result No. | Query Match | Score | Length | ID | Description | Pred. No. |
|------------|-------------|-------|--------|----|-------------|-----------|
|------------|-------------|-------|--------|----|-------------|-----------|

No matches found.

Search completed: Sat May 13 04:53:24 2000

Job time : 162 secs.

THIS PAGE BLANK



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Sat May 13 04:59:02 2000; Maspar time 3.04 Seconds
Tabular output not generated. 116.885 Million cell updates/sec

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pap
Perfect Score: 113
Sequence: 1 CECNIKVKVDNDNFP 15

Scoring table: PAM 150
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: a-geneseg35
1:genesegp

Statistics: Mean 18.468; Variance 59.453; scale 0.311
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|--------|----------------------------------|-----------|
| 1 | 113 | 100.0 | 15 | W04845 | Self epitope of desmog 3.35e-05 | |
| 2 | 52 | 46.0 | 11 | W13139 | Human cadherin-5 antiag 1.72e+02 | |

Note: Post-processor removed 43 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID W04845 standard; peptide: 15 AA.
AC W04845:
DE 18-FEB-1997 (first entry)
KW Self epitope of desmoglein 3, implicated in autoimmune disease.
KW Tolerisation; self-epitope; antigen; autoimmune disease;
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW pemphigus vulgaris; desmoglein; multiple sclerosis;
KW herpes simplex virus; adenovirus; phosphomannomutase;
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW Influenza; haemagglutinin; reovirus; sigma protein.
OS Homo sapiens.
PN W09627387-A1.
PD 12-SEP-1996.

PF 07-MAR-1996; U03182.
PA 07-MAR-1995; US-400796.
PR (HARD) HARVARD COLLEGE.
PI Strominger JL, Mucherpfennig KW;
DR WPI: 96-425218/42.
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT antigens - useful in disease treatment, and method for
PT identification of other self and non-self antigens implicated in
PT autoimmune disease
PS Claim 1; Page 40; 58pp; English.
CC Pharmaceutical preparations for tolerisation to antigens comprise
CC either an isolated human non-collagen or non-mysin basic protein
CC (MBP) polypeptide which is capable of tolerising an individual to an
CC autoantigen; or an isolated human pathogen polypeptide capable of
CC tolerising an individual to that polypeptide. In both cases, the
CC sequence corresponding to a sequence motif for a MHC class II
CC protein, such as HLA-DR, which is associated with a human autoimmune
CC disease, and which binds to the polypeptide to activate autoreactive
CC T-cells in individuals with the autoimmune disease. This peptide is
CC derived from the human desmoglein 3 protein (amino acids 251-285)
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC derived from the human desmoglein protein are described in W04841-47.
SQ Sequence 15 AA:

Query Match 100.0%; Score 113; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.35e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 CECNIKVKVDNDNFP 15
Oy 1 CECNIKVKVDNDNFP 15

RESULT 2
ID W13139 standard; Protein; 11 AA.
AC W13139:
DE 14-MAY-1997 (first entry)
KW Human cadherin-5 antigenic epitope (residues 242-252).
KW Ca2+ dependent; cell adhesion protein; cadherin; human; antibody;
KW purification; determination; epitope; tissue expression;
KW binding antagonist; calcium ion; antigen.
OS Homo sapiens.
PN US5597725-A.
PD 28-JAN-1997.
PF 17-APR-1992; 872643.
PR 17-APR-1992; US-872643.
PR 19-APR-1993; US-049460.
PR 26-JAN-1994; US-188228.
PA (DOHE-) DOHENY EYE INSTR.
PI Suzuki S:
DR WPI: 97-108328/10.
PT Antibodies to cadherin proteins - useful as cadherin antagonists,
PT etc.
PS Claim 5; Column 112; 59pp; English.
CC The present sequence is an antigenic epitope from human cadherin-5,
CC which is a Ca2+ dependent cell adhesion protein. Antibodies or
CC fragments that specifically bind the epitope can be used to purify
CC the cadherin, determine its tissue expression and antagonise its
CC ligand/antiligand binding activities.
SQ Sequence 11 AA:

Query Match 46.0%; Score 52; DB 1; Length 11;
Best Local Similarity 54.5%; Pred. No. 1.72e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 1 VTLODINDNFP 11
Oy 5 IKVKVDNDNFP 15

Search completed: Sat May 13 04:59:09 2000
Job time : 7 secs.

WORLDWIDE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:41:35 2000; Mspar time 406.97 Seconds
4.114 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pap
Perfect Score: 113
Sequence: 1 CECNIKVKVDNDNFP 15

Scoring table:
PAM 150
Gap 15

Searched: 714217 seqs, 111628056 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database:

a-pending
1:PCT 2:06 3:U60 4:07 5:U80 6:U81 7:U82 8:U83 9:U84A
10:U84B 11:U85 12:U86 13:U87 14:U88 15:U89 16:U90 17:U91
18:U92 19:U93 20:U94 21:U95 22:NEW0 23:NEW060 24:NEW08
25:NEW09

Statistics: Mean 20.820; Variance 56.137; scale 0.371

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

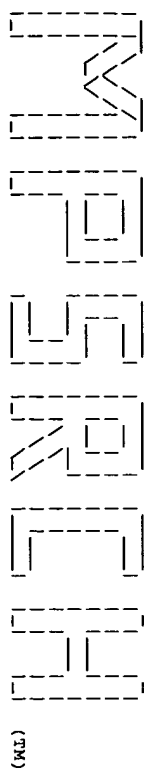
| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|------------|-------|-------------|--------|-------|-------------|---------------------------------|
| 1 | 113 | 100.0 | 15 | 1 | PCT-US96-0 | Sequence 5, Applicatio 1.48e-05 |
| 2 | 113 | 100.0 | 15 | 15 | US-08-991- | Sequence 5, Applicatio 1.48e-05 |

Note: Post-processor removed 43 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
STD PCT-US96-03182-5 STANDARD; PRT: 15 AA.
AC xxxxxx
DT
XX
DE Sequence 5, Application PC/TUS9603182
XX Sequence 5, Application PC/TUS9603182
CC GENERAL INFORMATION:
CC

CC APPLICANT:
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF
CC TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC STREET: 600 ATLANTIC AVENUE
CC CITY: BOSTON
CC STATE: MASSACHUSETTS
CC COUNTRY: UNITED STATES OF AMERICA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/03182
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/400,796
CC FILING DATE: 07-MAR-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: GATES, EDWARD R.
CC REGISTRATION NUMBER: 31,616
CC REFERENCE/DOCKET NUMBER: H0498/7015WO
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-3500
CC TELEFAX: 617-720-2441
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: internal
CC ORIGINAL SOURCE:
CC ORGANISM: HOMO SAPIENS
CC SEQUENCE 15 AA: 1738 MW; 1181 CN;
SQ
Query Match 100.0%; Score 113; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.48e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1 CECNIKVKVDNDNFP 15
QY 1 CECNIKVKVDNDNFP 15
RESULT 2
ID US-08-991-628-5 STANDARD; PRT: 15 AA.
AC xxxxxx
DT
XX
DE Sequence 5, Application US/08991628
XX
CC Sequence 5, Application US/08991628
CC GENERAL INFORMATION:
CC APPLICANT: STROMINGER, Jack L.
CC APPLICANT: WUCHERPFENNIG, Kai
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF ANTIGENS IMPLICATED
CC FILE REFERENCE: HAR-001DV
CC CURRENT APPLICATION NUMBER: US/08/991,628
CC CURRENT FILING DATE: 1997-11-15
CC EARLIER APPLICATION NUMBER: USSN 08/400,796
CC EARLIER FILING DATE: 1995-03-07
CC NUMBER OF SEQ ID NOS: 20
CC SOFTWARE: PatentIn Ver. 2.0
CC



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:58:36 2000; Mapar time 4.28 Seconds
165.342 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pep
Perfect Score: 113
Sequence: I CECNIKVKVDVNDNFP 15

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: plr62
1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 25.922; Variance 35.935; scale 0.721

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Length | ID | Description | Pred. No. |
|------------|-------------|--------|----|-------------|-----------|
|------------|-------------|--------|----|-------------|-----------|

No matches found.

Search completed: Sat May 13 04:58:44 2000
Job time : 8 secs.

 WIDEORIT
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:54:37 2000; MasPar time 79.85 Seconds
 5.721 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-5
 Description: (1-15) from US08991628.pep
 Perfect Score: 113
 Sequence: 1 CECNKKYKVDVNDNFP 15

Scoring table: PAM 150
 Gap 15

Searched: 83857 segs, 30454973 residues

Post-processing: Minimum Match 08
 Listing first 45 summaries
 Maximum DB seq length 15

Database: swiss-prot38
 1:swissprot

Statistics: Mean 26.676; Variance 31.455; scale 0.848

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Length | ID | Description | Pred. No. |
|-------------------|-------------|--------|----|-------------|-----------|
| No matches found. | | | | | |

Search completed: Sat May 13 04:56:06 2000
 Job time : 89 secs.

 M P E R C H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPerch_p protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:56:23 2000; MasPar time 105.85 Seconds
 9.825 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-5
 Description: (1-15) from US08991628.pep
 Perfect Score: 113
 Sequence: 1 CECNIRKVDVNDNFP 15

Scoring table: PAM 150
 Gap 15

Searched: 225878 segs, 69334122 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries
 Maximum DB seq length 15

Database: sptrembl12
 1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
 5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
 9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
 13:sp_vertebrate 14:sp_virus

Statistics: Mean 25.749; Variance 32.815; Scale 0.785

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

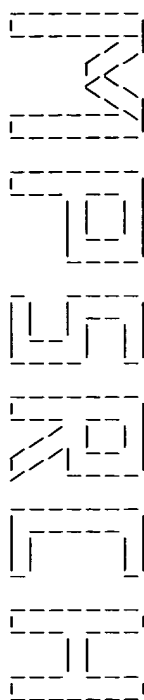
SUMMARIES

| Result No. | Query | Score | Match | Length | ID | Description | Pred. No. |
|------------|-------|-------|-------|--------|----|-------------|-----------|
| ----- | | | | | | | |

No matches found.

Search completed: Sat May 13 04:58:17 2000
 Job time : 114 secs.

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 05:06:08 2000; Maspar time 3.03 Seconds
117.247 Million cell updates/sec
Tabular output not generated.

Title: >US-08-991-628-6
Description: (1-15) from US08991628.pep
Perfect Score: 109
Sequence: 1 SARTLNRRYTGPTTF 15

Scoring table: PAM 150
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: a-geneseg35
1:genesegp

Statistics: Mean 19.232; Variance 58.683; scale 0.328
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | No. | Score | Match | Length | ID | Description | Pred. No. |
|--------|-----|-------|-------|--------|--------|---------------------------------|-----------|
| 1 | 109 | 100.0 | 15 | 1 | W04846 | Self epitope of desmog 5.72e-05 | |

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID W04846 standard; peptide: 15 AA.
AC W04846;
DT 18-FEB-1997 (first entry)
DE Self epitope of desmoglein 3, implicated in autoimmune disease.
KW Tolerisation; self-epitope; antigen; autoimmune disease;
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW pemphigus vulgaris; desmoglein; multiple sclerosis;
KW herpes simplex virus; adenovirus; phosphomannomutase;
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW Influenza; haemagglutinin; reovirus; sigma protein.
OS Homo sapiens.
PN W09627387-A1.
PD 12-SEP-1996.
PF 07-MAR-1996; U03182.

PR 07-MAR-1995; US-400796.
PA (HARD) HARVARD COLLEGE.
PI Strominger JL, Wucherpfennig KW;
DR WPI; 96-425218/42.
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT antigens - useful in disease treatment, and method for
PT identification of other self and non-self antigens implicated in
PT auto-immune disease
PS Claim 1; Page 41; 58pp; English.
CC Pharmaceutical preparations for tolerisation to antigens comprise
CC either an isolated human non-collagen or non-mysin basic protein
CC (MBP) polypeptide which is capable of tolerising an individual to an
CC autoantigen; or an isolated human pathogen polypeptide capable of
CC tolerising an individual to that polypeptide. In both cases, the
CC sequence corresponding to a sequence motif for a MHC class II
CC protein, such as HLA-DR, which is associated with a human autoimmune
CC disease and which binds to the polypeptide to activate autoreactive
CC T-cells in individuals with the autoimmune disease. This peptide is
CC derived from the human desmoglein 3 protein (amino acids 512-526)
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC derived from the human desmoglein protein are described in W04841-47.
SQ Sequence 15 AA;

Query Match 100.0%; Score 109; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 SARTLNRRYTGPTTF 15
QY 1 SARTLNRRYTGPTTF 15

Search completed: Sat May 13 05:06:15 2000
Job time : 7 secs.

M P E R I H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPearch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:49:00 2000; Maspar time 50.21 Seconds

3.871 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-6
Description: (1-15) from US08991628.pep
Perfect Score: 109
Sequence: 1 SARTLNRRYTGPTTF 15

Scoring table:
PAM 150
Gap 15

Searched: 131253 seqs, 12956647 residues

Post-Processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: a-1ssued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 18.236; Variance 59.295; scale 0.308

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

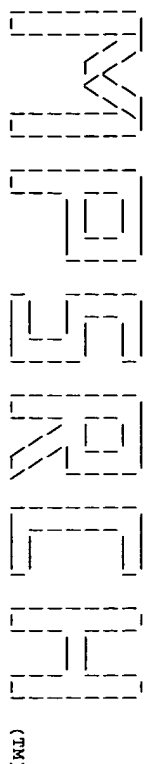
| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|--------------|------------------------|-----------|
| 1 | 109 | 100.0 | 15 | 2 US-08-400- | Sequence 6, Applicatio | 7.06e-05 |

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID US-08-400-796-6 STANDARD: PRT; 15 AA.
AC xxxxxx
XX
XX
DT
XX
XX
DE Sequence 6, Application US/08400796
XX
CC Patent No. 5874531
CC GENERAL INFORMATION:
CC APPLICANT: STROMINGER, JACK L.
CC APPLICANT: WUCHERPFENNIG, KAI
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF

CC TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC STREET: 600 ATLANTIC AVENUE
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/400.796
CC FILING DATE:
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: TWOMEY, MICHAEL J.
CC REGISTRATION NUMBER: 38,349
CC REFERENCE/DOCKET NUMBER: H0498/7015
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-3500
CC TELEFAX: 617-720-2441
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHEICAL: NO
CC FRAGMENT TYPE: internal
CC ORIGINAL SOURCE:
CC ORGANISM: HOMO SAPIENS
CC SEQUENCE 15 AA; 1761 MW; 1506 CN;
SQ
Query Match 100.0%; Score 109; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 7.06e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 SARTLNRRYTGPTTF 15
QY 1 SARTLNRRYTGPTTF 15
Search completed: Sat May 13 05:50:03 2000
Job time: 63 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 05:50:23 2000; Maspar time 402.79 seconds
4.157 million cell updates/sec
Tabular output not generated.

Title: >US-08-991-628-6
Description: (1-15) from US08991628.pep
Perfect Score: 109
Sequence: 1 SARTLNRRYTGPTTF 15

Scoring table:
PAM 150
Gap 15
Searched: 714217 seqs, 111628056 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database:
a-pending
1: PCT 2: US 3: US60 4: US7 5: US80 6: US81 7: US82 8: US83 9: US84
10: US84B 11: US85 12: US86 13: US87 14: US88 15: US89 16: US90 17: US91
18: US92 19: US93 20: US94 21: US95 22: NEWP 23: NEWU60 24: NEWU8
25: NEWU9

Statistics: Mean 21.442; Variance 54.363; scale 0.394

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|------------|-------|-------------|--------|-------|-------------|------------------------|
| 1 | 109 | 100.0 | 15 | 1 | PCT-US96-0 | Sequence 6, Applicatio |
| 2 | 109 | 100.0 | 15 | 15 | US-08-991- | Sequence 6, Applicatio |
| 3 | 56 | 51.4 | 10 | 19 | US-09-361- | Sequence 6, Applicatio |

Note: Post-processor removed 42 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID PCT-US96-03182-6 STANDARD; PRT: 15 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 6, Application PC/TUS9603182
XX Sequence 6, Application PC/TUS9603182
CC

GENERAL INFORMATION:

CC APPLICANT:
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF
CC TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC STREET: 600 ATLANTIC AVENUE
CC CITY: BOSTON
CC STATE: MASSACHUSETTS
CC COUNTRY: UNITED STATES OF AMERICA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/03182
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/400,796
CC FILING DATE: 07-MAR-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: GATES, EDWARD R.
CC REGISTRATION NUMBER: 31,616
CC REFERENCE/DOCKET NUMBER: H0498/7015WO
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-2441
CC TELEFAX: 617-720-3500
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: internal
CC ORIGIN: SOURCE: Internal
CC ORGANISM: HOMO SAPIENS
CC SEQ SEQUENCE 15 AA; 1761 MW; 1506 CN;

Query Match 100.0%; Score 109; DB 1; Length 15;

Best Local Similarity 100.0%; Pred. No. 2.15e-05; Mismatches 0; Indels 0; Gaps 0;

DB 1 SARTLNRRYTGPTTF 15
1 SARTLNRRYTGPTTF 15

RESULT 2
ID US-08-991-628-6 STANDARD; PRT: 15 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 6, Application US/08991628
XX
CC Sequence 6, Application US/08991628
CC GENERAL INFORMATION:
CC APPLICANT: STROMINGER, Jack L
CC APPLICANT: WUCHERPFENNIG, Kai
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF ANTIGENS IMPLICATED
CC FILE REFERENCE: HAR-001DV
CC CURRENT APPLICATION NUMBER: US/08/991,628
CC CURRENT FILING DATE: 1997-11-15
CC EARLIER APPLICATION NUMBER: USSN 08/400,796
CC EARLIER FILING DATE: 1995-03-07
CC NUMBER OF SEQ ID NOS: 20

WARNING (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_dp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 05:05:40 2000; MasPar time 4.33 Seconds
163.439 Million cell updates/sec
Tabular output not generated.

Title: >US-08-991-628-6
Description: (1-15) from US08991628.pep
Perfect Score: 109
Sequence: 1 SARTLNRRYTGPTTF 15
Scoring table: PAM 150
Gap 15

Searched: 142080 segs, 47172406 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 26.736; Variance 36.896; scale 0.725
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Score | Length | ID | Description | Pred. No. |
|-------------------|-------------|-------|--------|----|-------------|-----------|
| No matches found. | | | | | | |

Search completed: Sat May 13 05:05:49 2000
Job time : 9 secs.

 M P E R L H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:59:27 2000; Maspar time 83.07 Seconds
 5.499 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-6
 Description: (1-15) from US08991628.pep
 Perfect Score: 109
 Sequence: 1 SARTLNRRYTGPTTF 15

Scoring table:
 PAM 150
 Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries
 Maximum DB seq length 15

Database: swiss-prot38
 1:swissprot

Statistics: Mean 27.244; Variance 33.691; scale 0.809

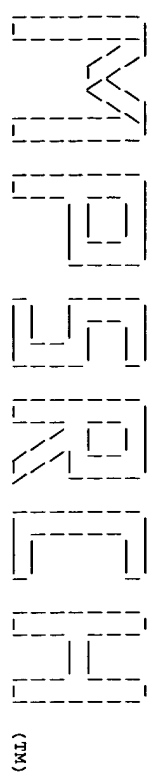
Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query | No. | Score | Match | Length | ID | Description | Pred. | No. |
|--------|-------|-----|-------|-------|--------|----|-------------|-------|-----|
| ----- | | | | | | | | | |

No matches found.

Search completed: Sat May 13 05:01:04 2000
 Job time : 97 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp .protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 05:01:23 2000; Maspar time 224.41 Seconds
Tabular output not generated. 4.634 Million cell updates/sec

Title: >US-08-991-628-6
Description: (1-15) from US08991628.pep
Perfect score: 109
Sequence: 1 SARTLNRRYTGPTTF 15

Scoring table:
PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_minc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_protent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 26.282; Variance 35.962; scale 0.731

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

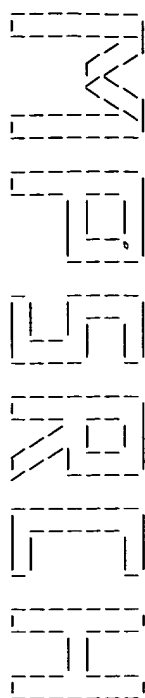
SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|-------|-------------|-----------|
| ----- | ----- | ----- | ----- | ----- | ----- | ----- |

No matches found.

Search completed: Sat May 13 05:05:22 2000
Job time : 239 secs.

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 05:13:41 2000; Maspar time 3.01 Seconds
118.170 Million cell updates/sec
Tabular output not generated.

Title: >US-08-991-628-7
Description: (1-15) From US08991628.pep
Perfect Score: 99
Sequence: 1 QSGTMRTRHSTGCTN 15

Scoring table:
PAM 150
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: a-geneseg35
1:genesegp

Statistics: Mean 17.183; Variance 54.742; scale 0.314

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Score | Match Length | DB ID | Description | Pred. No. |
|------------|-------------|--------------|-------|--|-----------|
| 1 | 99 | 100.0 | 15 | 1 W04847 Self epitope of desmog 5.70e-04 | |

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID W04847 standard; peptide: 15 AA.
AC W04847;
DE 18-FEB-1997 (first entry)
KW Self epitope of desmoglein 3, implicated in autoimmune disease.
KW Tolerisation; self-epitope; antigen; autoimmune disease;
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW pemphigus vulgaris; desmoglein; multiple sclerosis;
KW herpes simplex virus; adenovirus; phosphomannomutase;
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW influenza; haemagglutinin; reovirus; sigma protein.
OS Homo sapiens.
PN W09627387-A1.
PD 12-SEP-1996.
PF 07-MAR-1996; U03182.

PR 07-MAR-1995; US-400796.
PA (HARD) HARVARD COLLEGE.
PI Strominger JL, Muenchpfeinlg KW;
DR WPI: 96-425218/42.
PR Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT antigens - useful in disease treatment, and method for
PT identification of other self and non-self antigens implicated in
PR auto-immune disease.
PS Claim 1; Page 42; 56pp; English.
CC Pharmaceutical preparations for tolerisation to antigens comprise
CC either an isolated human non-collagen or non-mysin basic protein
CC (MBP) polypeptide which is capable of tolerising an individual to an
CC autoantigen; or an isolated human pathogen polypeptide capable of
CC tolerising an individual to that polypeptide. In both cases, the
CC polypeptide (whether self or non-self) includes an amino acid
CC sequence corresponding to a sequence motif for a MHC class II
CC protein, such as HLA-DR, which is associated with a human autoimmune
CC disease and which binds to the polypeptide to activate autoreactive
CC T-cells in individuals with the autoimmune disease. This peptide is
CC derived from the human desmoglein 3 protein (amino acids 763-786)
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC derived from the human desmoglein protein are described in W04841-47.
SQ Sequence 15 AA;

Query Match 100.0%; Score 99; DB 1; Length 15;
Best local similarity 100.0%; Pred. No. 5.70e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1 QSGTMRTRHSTGCTN 15
QY 1 QSGTMRTRHSTGCTN 15

Search completed: Sat May 13 05:13:48 2000
Job time : 7 secs.

Mon May 15 08:09:23 2000

US-08-991-628-7.rai

Page 1

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

MPsich_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 05:57:42 2000; MasPar time 47.18 Seconds
Tabular output not generated. 4.120 Million cell updates/sec

```

Title: >US-08-991-628-7
Description: (1-15) from US08991628.pep
Perfect Score: 99
Sequence: 1 QSGTMRTRHSTGCTN 15

Scoring table: PAM 150
Gap 15

Searched: 131253 seqs, 12956647 residues

```
Post-processing: Minimum Match 0%
                  Listing first 45 summaries
                  Maximum DB seq length 15
```

```
Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1
```

Statistics: Mean 16.273; Variance 51.318; scale 0.317

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|------------|------------------------|-----------|
| 1 | 99 | 100.0 | 15 | 2 | US-08-400- | Sequence 7, Applicatio | 1.88e-04 |

Note: Post-processor removed 44 summaries from 11st due to search parameters chosen.

ALIGNMENTS

| | | | |
|--------|---|-----------|-------------|
| RESULT | 1 | | |
| ID | US-08-400-796-7 | STANDARD; | PRT; 15 AA. |
| XX | | | |
| AC | xxxxxx | | |
| XX | | | |
| DT | | | |
| XX | | | |
| DE | Sequence 7, Application US/08400796 | | |
| XX | | | |
| CC | Sequence 7, Application US/08400796 | | |
| CC | Patent No. 5874531 | | |
| CC | GENERAL INFORMATION: | | |
| CC | APPLICANT: STROMINGER, JACK L. | | |
| CC | APPLICANT: WUCHERPFENNIG, KAI | | |
| CC | TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF | | |

```

CC      TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC      NUMBER OF SEQUENCES: 16
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC      STREET: 600 ATLANTIC AVENUE
CC      CITY: BOSTON
CC      STATE: MA
CC      COUNTRY: USA
CC      ZIP: 02210
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/400,796
CC      FILING DATE:
CC      CLASSIFICATION: 424
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: TWOMEY, MICHAEL J.
CC      REGISTRATION NUMBER: 38,349
CC      REFERENCE/DOCKET NUMBER: H0498/7015
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: 617-720-3500
CC      TELEFAX: 617-720-2441
CC      INFORMATION FOR SEQ ID NO: 7:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 15 amino acids
CC      type: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: peptide
CC      HYPOTHETICAL: NO
CC      FRAGMENT TYPE: Internal
CC      ORIGINAL SOURCE:
CC      ORGANISM: HOMO SAPIENS
CC      SEQUENCE: 15 AA; 1591 MW; 1253 CN;

Query Match      100.0%; Score 99; DB 2; Length 15;
Best Local Similarity 100.0%; Pred.No. 1.89e-04;
Matches      15; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Db      1 OSGMTRRHSTGCTN 15
      1 OSGMTRRHSTGCTN 15
      1 OSGMTRRHSTGCTN 15

```

Search completed: Sat May 13 05:58:40 2000
Job time : 58 secs.

M I O S R I H

(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:59:02 2000; Maspar time 398.09 Seconds
4.206 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-7
Description: (1-15) from US08991628.pep
Perfect Score: 99
Sequence: 1 QSGTMRTRHSTGCTN 15

Scoring table:
PAM 150
Gap 15

Searched: 714217 seqs, 111628056 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database:
a-pending
1:PCT 2:06 3:060 4:07 5:080 6:081 7:082 8:083 9:084A
10:084B 11:085 12:086 13:087 14:088 15:089 16:090 17:091
18:092 19:093 20:094 21:095 22:NEWP 23:NEWU60 24:NEWU8
25:NEWU9

Statistics: Mean 19.349; Variance 49.952; scale 0.387

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|---------------|-------|----------------|--------|----------|-------------|---------------------------------|
| 1 | 99 | 100.0 | 15 | 1 | PCT-US96-0 | Sequence 7, Applicatio 1.97e-04 |
| 2 | 99 | 100.0 | 15 | 15 | US-08-991- | Sequence 7, Applicatio 1.97e-04 |

Note: Post-processor removed 43 summaries from list due to search parameters chosen.

ALIGNMENTS

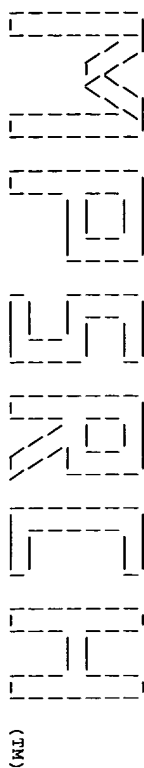
| RESULT | 1 | STANDARD; | PRT; | 15 | AA. |
|--------|---------------------------------------|-----------|------|----|-----|
| ID | PCT-US96-03182-7 | | | | |
| XX | xxxxxx | | | | |
| AC | | | | | |
| XX | | | | | |
| DT | | | | | |
| XX | | | | | |
| DE | Sequence 7, Application PC/TUS9603182 | | | | |
| XX | | | | | |
| CC | Sequence 7, Application PC/TUS9603182 | | | | |
| CC | GENERAL INFORMATION: | | | | |

| | | | | | |
|--------|---|-----------|------|----|-----|
| CC | APPLICANT: | | | | |
| CC | TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF | | | | |
| CC | TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE | | | | |
| CC | NUMBER OF SEQUENCES: 16 | | | | |
| CC | CORRESPONDENCE ADDRESSES: | | | | |
| CC | ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C. | | | | |
| CC | STREET: 600 ATLANTIC AVENUE | | | | |
| CC | CITY: BOSTON | | | | |
| CC | STATE: MASSACHUSETTS | | | | |
| CC | COUNTRY: UNITED STATES OF AMERICA | | | | |
| CC | ZIP: 02210 | | | | |
| CC | COMPUTER READABLE FORM: | | | | |
| CC | MEDIUM TYPE: Floppy disk | | | | |
| CC | COMPUTER: IBM PC compatible | | | | |
| CC | OPERATING SYSTEM: PC-DOS/MS-DOS | | | | |
| CC | SOFTWARE: PatentIn Release #1.0, Version #1.25 | | | | |
| CC | CURRENT APPLICATION DATA: | | | | |
| CC | APPLICATION NUMBER: PCT/US96/03182 | | | | |
| CC | FILING DATE: | | | | |
| CC | CLASSIFICATION: | | | | |
| CC | PRIOR APPLICATION DATA: | | | | |
| CC | APPLICATION NUMBER: US 08/400,796 | | | | |
| CC | FILING DATE: 07-MAR-1995 | | | | |
| CC | ATTORNEY/AGENT INFORMATION: | | | | |
| CC | NAME: GATES, EDWARD R. | | | | |
| CC | REGISTRATION NUMBER: 31,616 | | | | |
| CC | REFERENCE/DOCKET NUMBER: H0496/7015MO | | | | |
| CC | TELECOMMUNICATION INFORMATION: | | | | |
| CC | TELEPHONE: 617-720-3500 | | | | |
| CC | TELEFAX: 617-720-2441 | | | | |
| CC | INFORMATION FOR SEQ ID NO: 7: | | | | |
| CC | SEQUENCE CHARACTERISTICS: | | | | |
| CC | LENGTH: 15 amino acids | | | | |
| CC | TYPE: amino acid | | | | |
| CC | TOPOLOGY: linear | | | | |
| CC | MOLECULE TYPE: peptide | | | | |
| CC | HYPOTHETICAL: NO | | | | |
| CC | FRAGMENT TYPE: Internal | | | | |
| CC | ORIGINAL SOURCE: | | | | |
| CC | ORGANISM: HOMO SAPIENS | | | | |
| CC | SEQUENCE 15 AA: 1591 MM: 1253 CN: | | | | |
| CC | Query Match 100.0%; Score 99; DB 1; Length 15; | | | | |
| CC | Best Local Similarity 100.0%; Pred. No. 1.97e-04; | | | | |
| CC | Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | | | |
| DB | 1 QSGTMRTRHSTGCTN 15 | | | | |
| QY | 1 QSGTMRTRHSTGCTN 15 | | | | |
| RESULT | 2 | STANDARD; | PRT; | 15 | AA. |
| ID | US-08-991-628-7 | | | | |
| XX | xxxxxx | | | | |
| AC | | | | | |
| XX | | | | | |
| DT | | | | | |
| XX | | | | | |
| DE | Sequence 7, Application US/08991628 | | | | |
| XX | | | | | |
| CC | Sequence 7, Application US/08991628 | | | | |
| CC | GENERAL INFORMATION: | | | | |
| CC | APPLICANT: STROMINGER, Jack L | | | | |
| CC | APPLICANT: WUCHERPRENNIG, Kai | | | | |
| CC | TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF ANTIGENS IMPLICATED | | | | |
| CC | TITLE OF INVENTION: IN AUTOIMMUNE DISEASE | | | | |
| CC | FILE REFERENCE: HAR-001DV | | | | |
| CC | CURRENT APPLICATION NUMBER: US/08/991,628 | | | | |
| CC | CURRENT FILING DATE: 1997-11-15 | | | | |
| CC | EARLIER APPLICATION NUMBER: USSN 08/400,796 | | | | |
| CC | EARLIER FILING DATE: 1995-03-07 | | | | |
| CC | NUMBER OF SEQ ID NOS: 20 | | | | |
| CC | SOFTWARE: PatentIn Ver. 2.0 | | | | |

CC SEQ ID NO 7
 CC LENGTH: 15
 CC TYPE: PRT
 CC ORGANISM: Homo sapiens
 SQ SEQUENCE 15 AA: 1591 MW: 1253 CN:

Query Match 100.0%; Score 99; DB 15; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.97e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 OSGTMRTRHSTGCTN 15
 OY 1 OSGTMRTRHSTGCTN 15

Search completed: Sat May 13 06:05:57 2000
 Job time : 415 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:13:15 2000; MasPar time 4.30 Seconds
164.392 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-7
Description: (1-15) from US08991628.pep
Perfect Score: 99
Sequence: 1 QSGTMRTRHSTGCTN 15

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

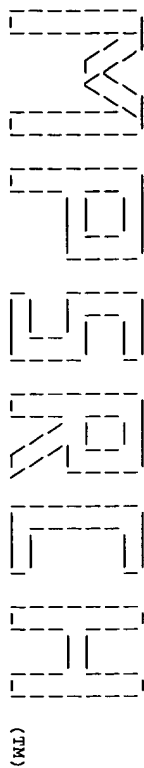
Statistics: Mean 23.943; Variance 29.203; scale 0.820

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Score | Match Length | ID | Description | Pred. No. |
|-------------------|-------------|--------------|----|-------------|-----------|
| No matches found. | | | | | |

Search completed: Sat May 13 05:13:24 2000
Job time : 9 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_bp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:06:34 2000; MasPar time 91.30 Seconds
5.004 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-7
Description: (1-15) from US08991628.pep
Perfect Score: 99
Sequence: 1 QSGTMRTRHSTGCTN 15

Scoring table: PAM 150
Gap 15

Searched: 83857 segs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: swiss-prot38
1:swissprot

Statistics: Mean 24.519; Variance 27.183; scale 0.902

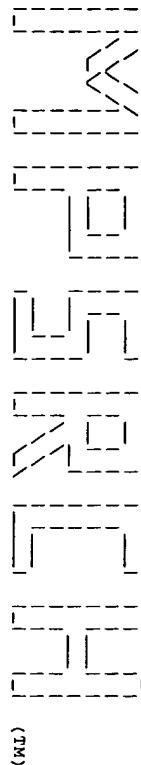
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Score | Match Length | ID | Description | Pred. No. |
|------------|-------------|--------------|----|-------------|-----------|
|------------|-------------|--------------|----|-------------|-----------|

No matches found.

Search completed: Sat May 13 05:08:16 2000
Job time : 102 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run On: Sat May 13 05:08:36 2000; MasPar time 245.94 Seconds
4.229 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-7
Description: (1-15) from US08991628.pep
Perfect Score: 99
Sequence: 1 QSGTMRTRHSTGCTN 15

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: splemb112
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 24.224; Variance 25.706; scale 0.942

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query | | | | |
|--------|-------|-------|--------|----|-------------|
| No. | Score | Match | Length | ID | Description |
| | | | | | Pred. No. |

No matches found.

Search completed: Sat May 13 05:12:56 2000
Job time : 260 secs.

 W0842
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 07:05:15 2000: Maspar time 3.02 Seconds
 Tabular output not generated. 117.657 Million cell updates/sec

Title: >US-08-991-628-2
 Description: (1-15) from US08991628.pep
 Perfect Score: 105
 Sequence: 1 FGIFVVDKNTGDINI 15

Scoring table: PAM 150
 Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-geneseq35
 1:geneseqp

Statistics: Mean 18.538; Variance 60.235; scale 0.308

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|--------|-------------------------|-----------|
| 1 | 105 | 100.0 | 15 | W0842 | Self epitope of desmog | 3.51e-04 |
| 2 | 105 | 100.0 | 30 | R93960 | Peptide combining with | 3.51e-04 |
| 3 | 105 | 100.0 | 614 | W07908 | Pemphigus vulgaris ant | 3.51e-04 |
| 4 | 105 | 100.0 | 999 | R30742 | Human pemphigus vulgar | 3.51e-04 |
| 5 | 86 | 81.9 | 19 | R93961 | Peptide combining with | 4.92e-02 |
| 6 | 86 | 81.9 | 21 | R93962 | Peptide combining with | 4.92e-02 |
| 7 | 84 | 80.0 | 778 | W15489 | Pemphigus foliaceus an | 8.20e-02 |
| 8 | 82 | 78.1 | 778 | W13009 | Segment of desmog | 1.36e-01 |
| 9 | 57 | 54.3 | 187 | Y11097 | H. pylori ORF hp7e1043 | 5.92e-01 |
| 10 | 57 | 54.3 | 204 | Y11098 | H. pylori ORF hp7e1043 | 5.92e-01 |
| 11 | 57 | 54.3 | 204 | W98278 | H. pylori GPO 876 pro | 5.92e-01 |
| 12 | 57 | 54.3 | 693 | R49732 | Sequence encoded by hu | 5.92e-01 |
| 13 | 57 | 54.3 | 796 | R49731 | Sequence encoded by hu | 5.92e-01 |
| 14 | 57 | 54.3 | 796 | W25636 | Human cadherin-11 | 5.92e-01 |
| 15 | 57 | 54.3 | 796 | R49730 | Sequence encoded by mu | 5.92e-01 |
| 16 | 57 | 54.3 | 796 | W85598 | Cadherin-11 | 5.92e-01 |
| 17 | 57 | 54.3 | 796 | W13134 | Full length human cadh | 5.92e-01 |
| 18 | 57 | 54.3 | 797 | R86866 | Human protocadherin pc | 5.92e-01 |
| 19 | 57 | 54.3 | 1400 | W82791 | Human RON receptor pro | 7.45e-01 |
| 20 | 56 | 53.3 | 433 | W41732 | Arabidopsis chloroplast | 7.45e-01 |
| 21 | 56 | 53.3 | 511 | R69506 | Aspergillus sp. recomb | 7.45e-01 |
| 22 | 56 | 53.3 | 515 | R69505 | Aspergillus sp. recomb | 7.45e-01 |
| 23 | 56 | 53.3 | 515 | R69504 | Aspergillus sp. recomb | 7.45e-01 |

| ID | Score | Query Match | Length | ID | Description | Pred. No. |
|----|-------|-------------|--------|--------|------------------------|-----------|
| 24 | 56 | 53.3 | 713 | W25638 | Human cadherin-13. | 7.45e-01 |
| 25 | 56 | 53.3 | 713 | W13136 | Full length human cadh | 7.45e-01 |
| 26 | 55 | 52.4 | 17 | R93963 | Peptide combining with | 9.36e-01 |
| 27 | 55 | 52.4 | 20 | W4250 | Human wild-type E-Cadn | 9.36e-01 |
| 28 | 55 | 52.4 | 499 | R60609 | Tobamovirus replicatio | 9.36e-01 |
| 29 | 55 | 52.4 | 878 | R55060 | Sequence of human live | 9.36e-01 |
| 30 | 55 | 52.4 | 878 | R85487 | Human E-cadherin precu | 9.36e-01 |
| 31 | 55 | 52.4 | 1827 | W74090 | Human hsi protein sequ | 9.36e-01 |
| 32 | 54 | 51.4 | 246 | R25789 | Sequence encoded by ch | 1.18e-02 |
| 33 | 54 | 51.4 | 246 | W76815 | N. gonorrhoeae pilQ pr | 1.18e-02 |
| 34 | 54 | 51.4 | 557 | W43448 | Tobacco laccase clone | 1.18e-02 |
| 35 | 53 | 50.5 | 199 | R34398 | Helicobacter pylori ur | 1.47e-02 |
| 36 | 53 | 50.5 | 794 | W25637 | Human cadherin-12. | 1.47e-02 |
| 37 | 53 | 50.5 | 794 | W13135 | Putative human cadheri | 1.47e-02 |
| 38 | 53 | 50.5 | 2647 | W13349 | Human filamin. | 1.47e-02 |
| 39 | 52 | 49.5 | 115 | W62680 | Streptococcus pneumoni | 1.84e-02 |
| 40 | 52 | 49.5 | 507 | W29772 | Malassezia fungus MF-5 | 1.84e-02 |
| 41 | 52 | 49.5 | 535 | R56549 | Cold acclimatization p | 1.84e-02 |
| 42 | 52 | 49.5 | 1026 | R58906 | Human protocadherin-42 | 1.84e-02 |
| 43 | 52 | 49.5 | 1026 | R87146 | Product of alternative | 1.84e-02 |
| 44 | 52 | 49.5 | 1203 | R58911 | Product of alternative | 1.84e-02 |
| 45 | 52 | 49.5 | 1203 | R87152 | Alternatively spliced | 1.84e-02 |

ALIGNMENTS

RESULT 1
 ID W0842 standard; peptide: 15 AA.
 AC W0842:
 DT 18-FEB-1997 (first entry)
 DE Self epitope of desmoglein 3, implicated in autoimmune disease.
 KW Tolerisation; self-epitope; antigen; autoimmune disease;
 KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
 KW pemphigus vulgaris; desmoglein; multiple sclerosis;
 KW herpes simplex virus; adenovirus; phosphonemovulase;
 KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
 KW influenza; haemagglutinin; reovirus; sigma protein.
 OS Homo sapiens.
 PN W09627387-A1.
 PD 12-SEP-1996.
 PF 07-MAR-1996; U03182.
 PR 07-MAR-1995; US-400796.
 PA (HARD) HARVARD COLLEGE.
 PI Strominger Jr., Muecherflemig KW:
 DR WPT: 96-425218/42.
 PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
 PT antigens - useful in disease treatment, and method for
 PT identification of other self and non-self antigens implicated in
 PT auto-immune disease
 PS Claim 1: Page 38: 58pp: English.
 CC Pharmaceutical preparations for tolerisation to antigens comprise
 CC either an isolated human non-collagen or non-mysin basic protein
 CC (MBP) polypeptide which is capable of tolerising an individual to an
 CC autoantigen; or an isolated human pathogen polypeptide capable of
 CC tolerising an individual to that polypeptide. In both cases, the
 CC polypeptide (whether self or non-self) includes an amino acid
 CC sequence corresponding to a sequence motif for a MHC class II
 CC protein, such as HLA-DR, which is associated with a human autoimmune
 CC disease and which binds to the polypeptide to activate autoreactive
 CC T-cells in individuals with the autoimmune disease. This peptide is
 CC derived from the human desmoglein 3 protein (amino acids 97-111)
 CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
 CC derived from the human desmoglein protein are described in W04841-47.
 SO Sequence 15 AA:

Query Match 100.0%; Score 105; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.51e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FGIFVVDKNTGDINI 15
 QY 1 FGIFVVDKNTGDINI 15

RESULT 2

ID R93960 standard: peptide: 30 AA.

AC R93960: (first entry)

DT 05-JUL-1996

DE Peptide combining with anti-interferidermal cellular antibody.

KW anti-interferidermal cellular antibody; autoantibody; adsorbent.

OS Synthetic.

PN J07309893-A.

PD 28-NOV-1995.

PF 18-MAY-1994: 129556.

PR 18-MAY-1994: JP-129556.

PA (KURS) KURARAY CO LTD.

DR WPI: 96-045392/05.

PT Anti-interferidermal cellular antibody-combining peptide - which can be immobilised on column to form adsorbent useful for treating diseases related to the antibody

PS Claim 1: Page 2: 7pp: Japanese.

CC A new peptide is disclosed which contains at least 5 contiguous amino acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe Val Val Asp Lys Asn Thr Gly Asp Ile Asp Thr Ala Ile Val Asp Arg Glu Glu (the present sequence), the peptide not containing more than 50 residues. The peptide combines with anti-interferidermal cellular antibody. It can be immobilised on a carrier to prepare an adsorbent useful for the treatment of diseases related to anti-interferidermal cellular antibody.

SO Sequence 30 AA:

Query Match 100.0%; Score 105; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.51e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 8 FGIFVVDKNTGDINI 22
QY 1 FGIFVVDKNTGDINI 15

RESULT 3

ID W07908 standard: protein: 614 AA.

AC W07908: (first entry)

DT 29-JAN-1997

DE Pemphigus vulgaris antigen protein extracellular region.

KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis; treatment; pemphigus vulgaris; PV; bulla; blister; skin disease; dermatology.

OS Homo sapiens.

PN J08188540-A.

PD 23-JUL-1996.

PF 30-JUN-1995: 165632.

PR 30-JUN-1994: JP-173291.

PA (NISH/) NISHUKAWA T.

DR WPI: 96-388562/39.

PT Fused protein recognised by pemphigus vulgaris autoantibody - useful to treat and diagnose pemphigus vulgaris

PS Claim 1: Page 7-9: 9pp: Japanese.

CC W07908 represents the human pemphigus vulgaris (PV) antigen extracellular region. The PV antigen is produced in patients with pemphigus vulgaris resulting in autoimmune disease. PV is a rare relapsing disease causing suprabasal, intra-epidermal bullae (vesicles) of the skin and mucous membranes, which is fatal if untreated. The PV antigen was fused to a human IgG1 hinge region and the resulting fusion protein is useful to treat or diagnose pemphigus vulgaris.

SO Sequence 614 AA:

Query Match 100.0%; Score 105; DB 1; Length 614;
Best Local Similarity 100.0%; Pred. No. 3.51e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 96 FGIFVVDKNTGDINI 110
QY 1 FGIFVVDKNTGDINI 15

RESULT 4

ID R30742 standard: protein: 999 AA.

AC R30742:

DT 14-JUN-1993 (first entry)

DE Human pemphigus vulgaris 130KD antigen.

KW Pemphigus vulgaris; skin disease; autoantibodies; keratinocyte cell surface antigen; glycoprotein; cell adhesion.

OS Homo sapiens.

PN US7798918-A.

PD 15-DEC-1992.

PF 27-NOV-1991: 798918.

PR 27-NOV-1991: US-798918.

PA (USSH) US DEPT HEALTH & HUMAN SERVICE.

PI Amagai M, Klaus-Kovtun V, Stanley JR;

DR WPI: 93-067436/08.

DR N-PEDA: Q35992.

PT DNA encoding pemphigus vulgaris antigen - useful in proteins for diagnostic and therapeutic uses

PS Disclosure: fig 7: 50pp: English.

CC This sequence is the pemphigus vulgaris 130KD antigen. The protein and its encoding DNA may be used in the diagnosis and treatment of pemphigus vulgaris. It is thought that the antigen may be a cell adhesion molecule.

SO Sequence 999 AA:

Query Match 100.0%; Score 105; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 3.51e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 97 FGIFVVDKNTGDINI 111
QY 1 FGIFVVDKNTGDINI 15

RESULT 5

ID R93961 standard: peptide: 19 AA.

AC R93961:

DT 05-JUL-1996 (first entry)

DE Peptide combining with anti-interferidermal cellular antibody.

KW anti-interferidermal cellular antibody; autoantibody; adsorbent.

OS Synthetic.

PN J07309893-A.

PD 28-NOV-1995.

PF 18-MAY-1994: 129556.

PR 18-MAY-1994: JP-129556.

PA (KURS) KURARAY CO LTD.

DR WPI: 96-045392/05.

PT Anti-interferidermal cellular antibody-combining peptide - which can be immobilised on column to form adsorbent useful for treating diseases related to the antibody

PS Example 1: Page 4: 7pp: Japanese.

CC New peptides are disclosed which contain at least 5 contiguous amino acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe Val Val Asp Lys Asn Thr Gly Asp Ile Asp Thr Ala Ile Val Asp Arg Glu Glu (see R93960), the peptide not containing more than 50 residues. The peptide combines with anti-interferidermal cellular antibody. It can be immobilised on a carrier to prepare an adsorbent useful for the treatment of diseases related to anti-interferidermal cellular antibody. The present sequence is a specific example of the new peptides.

SO Sequence 19 AA:

Query Match 81.9%; Score 86; DB 1; Length 19;
Best Local Similarity 100.0%; Pred. No. 4.92e-02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 8 FGIFVVDKNTGD 19
QY 1 FGIFVVDKNTGD 12

RESULT 6

ID R93962 standard: peptide: 21 AA.

AC R93962;
DT 05-JUL-1995 (first entry)
DE Peptide combining with anti-interferidermal cellular antibody.
KW anti-interferidermal cellular antibody; autoantibody; adsorbent.
OS Synthetic.
PN J07309893-A.
PD 28-NOV-1995.
PF 18-MAY-1994: 129556.
PR 18-MAY-1994: JP-129556.
PA (KURS.) KURARAY CO LTD.
WP: 96-045392/05.
DR Anti-interferidermal cellular antibody-combining peptide - which can be immobilised on column to form adsorbent useful for treating diseases related to the antibody.
PT Example 2: Page 5: 7pp. Japanese.
PS New peptides are disclosed which contain at least 5 contiguous amino acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe Val Val Asp Lys Asn Thr Gly Asp Ile Asn Ile Thr Ala Ile Val Asp Arg Glu Glu (see R93960), the peptide not containing more than 50 residues. The peptide combines with anti-interferidermal cellular antibody. It can be immobilised on a carrier to prepare an adsorbent useful for the treatment of diseases related to anti-interferidermal cellular antibody. The present sequence is a specific example of the new peptides.
SQ Sequence 21 AA;

Query Match 81.9%; Score 86; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.92e-02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 10 FGIFVNDKNTGD 21
OY 1 FGIFVNDKNTGD 12

RESULT 7
ID W15489 standard; Protein: 778 AA.
AC W15489;
DT 17-JUN-1997 (first entry)
DE Pemphigus foliaceus antigen-IgG constant region fusion protein.
KW Pemphigus foliaceus; autoantibody; constant region; IgG;
KW extracellular region; antigen; hinge portion; skin;
KW dermatitis herpetiformis; fusion protein; detection; ss.
OS Chimeric - Homo sapiens.
FS Key Location/Qualifiers
FT domain 1..545
PN J09077800-A.
PD 25-MAR-1997.
PF 12-SEP-1995: 260899.
PR 12-SEP-1995: JP-260899.
PA (NISHU.) NISHIKAWA T.
DR WP: 97-241758/22.
P-PSDB: T66428.
PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked through the hinge region used to treat pemphigus foliaceus
PS Claim 1: Page 10-12: 17pp; Japanese.
CC This sequence represents a fused protein recognised by pemphigus foliaceus patient autoantibody which comprises the constant region of IgG linked to the extracellular region of pemphigus foliaceus antigen protein through the hinge portion. Pemphigus foliaceus is a chronic, generalised, vesicular and scaling skin eruption similar to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion protein is useful to treat pemphigus foliaceus. The antigen is especially administered through an adsorbent upon which the fusion protein is immobilised via a carrier. The fusion protein is also useful for detecting pemphigus foliaceus antibodies which is useful in immunodiagnosis. The fusion protein has little or no side effects.
SQ Sequence 778 AA;

Query Match 80.0%; Score 84; DB 1; Length 778;
Best Local Similarity 60.0%; Pred. No. 8.20e-02;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 XGIFVNDKTEINI 111
OY 1 FGIFVNDKNTGDINI 15

RESULT 8
ID W13009 standard; Protein: 560 AA.
AC W13009;
DT 21-NOV-1997 (first entry)
DE Segment of desmosomal cadherin, desmoglein Dsg2.
KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW micrometastasis; separation; enrichment; targeted delivery;
KW metastatic.
OS Homo sapiens.
PN DE19531033-A1.
PD 27-FEB-1997.
PF 23-AUG-1995: 031033.
PR 23-AUG-1995: DE-031033.
PA (PROG-) PROGEN BIOTECHNIK GMBH.
PI Franke WW, Schaefer S:
WP: 97-146518/14.
DR Antibody reactive with part of desmosomal cadherin - exposed on surface of epithelial or carcinoma cells, not bound to desmosomes, useful for diagnosis and treatment of carcinoma micrometastases
PS Claim 7: Page 5: 8pp; German.
CC The present sequence is a segment of the desmosomal cadherin (DC), desmoglein Dsg2, which is exposed on the surface of epithelial or carcinoma cells and not bound to desmosomes. An antibody (Ab) directed against epitopes of the present sequence can be used to diagnose, i.e. to detect carcinoma cells, especially micrometastases, not bound to desmosomes, to separate, enrich or detect living or fixed carcinoma cells by cell sorting methods and as a therapeutic to deliver agents, e.g. other Ab or toxins, to target cells. The Ab provides rapid and reliable detection of metastatic carcinoma, and detects parts of DC that are not accessible in desmosome bound cells, as in normal tissue or carcinoma.
SQ Sequence 560 AA;

Query Match 78.1%; Score 82; DB 1; Length 560;
Best Local Similarity 60.0%; Pred. No. 1.36e-01;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 50 FGIFVNDKTEINLV 64
OY 1 FGIFVNDKNTGDINI 15

RESULT 9
ID Y11097 standard; Protein: 187 AA.
AC Y11097;
DT 08-JUN-1999 (first entry)
DE H. pylori ORF hp/elo433_36339535_f3.3 secreted protein.
KW Vaccines; probe; diagnostic; ORF; cell envelope protein.
KW secreted protein; cytoplasmic protein; cellular protein.
OS Helicobacter pylori.
PN M09824475-A1.
PD 11-JUN-1998.
PF 05-DEC-1997: U22104.
PR 14-JUL-1997: US-891928.
PR 05-DEC-1996: US-759625.
PR 25-MAR-1997: US-823745.
PA (ASPR.) ASTRA AB.
PI Alm RA, Castriotta LM, Dolg PC, Kabok Z, Smith D:
WP: 96-335011/29.
N-PSDB: X30626.
DR New isolated Helicobacter pylori nucleic acids - used to develop products for the diagnosis, prevention and treatment of infection by H. pylori and other Helicobacter species
PS Claims 37, 41: Page 264-265: 339pp; English.
CC Recombinant or substantially pure preparations of H. pylori polypeptides are disclosed, together with the nucleic acids encoding them. In all,

CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.
 CC Vaccines containing the nucleic acids or proteins are claimed, as are
 CC probes containing at least 8 nucleotides from the nucleic acid
 CC sequences. The vaccines are useful for treating or reducing the risk of
 CC H. pylori infections, and the probes can be used diagnostically for
 CC detecting the presence of Helicobacter in a sample. The products are
 CC also of use in screening for compounds having the ability to interfere
 CC with the H. pylori life cycle or to inhibit H. pylori infection.
 SO Sequence 187 AA.

Query Match 54.3%; Score 57; DB 1; Length 187;
 Best Local Similarity 46.2%; Pred. No. 5.92e+01;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

DB 67 GFFVVKKSANEID 79
 Oy 2 GFFVVDKNTGIDIN 14

RESULT 10
 ID Y11098 standard; Protein: 204 AA.

AC Y11098:1999 (first entry)
 DE H. pylori ORF hp7el0433_36339535_53_3 secreted protein.
 KW Vaccine: probe; diagnostic; ORF; cell envelope protein;
 KM secreted protein; cytoplasmic protein; cellular protein.
 OS Helicobacter pylori.
 PN MO9824475-A1.
 PD 11-JUN-1998.
 PF 05-DEC-1997; U22104.
 PR 14-JUL-1997; US-891928.
 PR 05-DEC-1996; US-759625.
 PR 25-MAR-1997; US-823745.
 PA (ASPR) ASTRA AB.
 PI Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;
 DR WPI: 96-333051/29.
 DR N-PSDB: X10627.
 PT New isolated Helicobacter pylori nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of infection by
 PT H. pylori and other Helicobacter species
 PS Claims 37, 41: Page 265-266: 339pp; English.
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.
 CC Vaccines containing the nucleic acids or proteins are claimed, as are
 CC probes containing at least 8 nucleotides from the nucleic acid
 CC sequences. The vaccines are useful for treating or reducing the risk of
 CC H. pylori infections, and the probes can be used diagnostically for
 CC detecting the presence of Helicobacter in a sample. The products are
 CC also of use in screening for compounds having the ability to interfere
 CC with the H. pylori life cycle or to inhibit H. pylori infection.
 SO Sequence 204 AA;

Query Match 54.3%; Score 57; DB 1; Length 204;
 Best Local Similarity 46.2%; Pred. No. 5.92e+01;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

DB 62 GFFVVKKSANEID 74
 Oy 2 GFFVVDKNTGIDIN 14

RESULT 11
 ID W98278 standard; Protein: 204 AA.

AC W98278:
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHPD 876 protein.
 KM GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease.
 OS Helicobacter pylori.
 PN MO9843478-A1.

PD 08-OCT-1998.
 PF 01-APR-1998; U06371.
 PR 29-JUL-1997; US-902615.
 PR 01-APR-1997; US-833457.
 PR 24-JUN-1997; US-881227.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 DR Al-Gazawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
 DR WPI: 98-542293/46.
 DR N-PSDB: X13997.
 PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 PS Claim 8: Page 315-316: 2054pp; English.
 CC This sequence represents a Helicobacter pylori GHPD protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 SO Sequence 204 AA;

Query Match 54.3%; Score 57; DB 1; Length 204;
 Best Local Similarity 46.2%; Pred. No. 5.92e+01;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

DB 62 GFFVVKKSANEID 74
 Oy 2 GFFVVDKNTGIDIN 14

RESULT 12
 ID R49732 standard; Protein: 693 AA.
 AC R49732:
 DT 14-SEP-1994 (first entry)
 DE Sequence encoded by human OSF-4-2 CDNA.
 KW OSF-4-2; cadherin; growth factor; osteogenesis; osteoblast; therapy;
 KM diagnosis.
 OS Homo sapiens.
 PN EP-585801-A.
 PD 09-MAR-1994.
 PF 25-AUG-1993; 113602.
 PR 28-AUG-1992; JP-230028.
 PA (FAHR) HOECHST JAPAN LTD.
 PI Aman N, Kawai S, Okazaki M, Takeshita S, Tsujimura A;
 DR WPI: 94-076152/10.
 DR N-PSDB: Q44393.
 PT New bone related, cadherin-like OSF-4 proteins - for treatment
 PT and diagnosis of bone metabolic disease, and nucleic acid
 PT encoding them
 PS Claim 1: Page 23-27: 34pp; English.
 CC cDNA libraries were constructed from the mouse osteoblastic cell
 CC line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
 CC then as much common DNA as possible removed by hybridisation between
 CC the 2 libraries. Residual E1-specific DNA was amplified. Inserted
 CC into lambda g10 and screened by plaque hybridisation. A minilink of
 CC 273 E-specific clones was recovered, their inserts amplified and
 CC used to screen total RNA from both cell types. One clone specific
 CC for E1 was identified and sequenced. The insert from this clone was
 CC used to screen cDNA prep. from E1 RNA and the longest posn. insert
 CC cloned in pGEM 11zf (+) to give pKOR164. This insert was sequenced;
 CC it encoded the 796 AA mouse precursor protein (Q44391/R49730). The
 CC insert was also used to screen a cDNA bank prep. from human
 CC osteosarcoma to identify 2 clones encoding the 2 human precursor
 CC proteins - OSF-4-1 and OSF-4-2 (Q44392/R49731 and Q44393/R49732
 CC respectively).

SO Sequence 693 AA;

Query Match 54.3%; Score 57; DB 1; Length 693;
 Best Local Similarity 50.0%; Pred. No. 5.92e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 100 IFVIDKSGN1H 111
 |||:|::|:|:
 Oy 3 IFVVDKNTGDIN 14

RESULT 13
 ID R49731 standard: Protein: 796 AA.
 AC R49731:
 DT 14-SEP-1994 (first entry)
 DE Sequence encoded by human OSF-4-1 cDNA.
 KW OSF-4-1: cadherin; growth factor; osteogenesis; osteoblast; therapy;
 KM diagnosis.
 OS Homo sapiens.
 PN EP-585801-A.
 PD 09-MAR-1994.
 PF 25-AUG-1993: 113602.
 PR 28-AUG-1992: JP-230028.
 PA (FARH) HOECHST JAPAN LTD.
 PI Amann E, Kawai S, Okazaki M, Takeshita S, Tsujimura A;
 DR N-PSDB: 044392/
 DR N-PSDB: 044392/
 PT New bone related, cadherin-like OSF-4 proteins - for treatment
 and diagnosis of bone metabolic disease, and nucleic acid
 encoding them
 PS Claim 1: Page 18-22: 34pp: English.
 CC cDNA libraries were constructed from the mouse osteoblastic cell
 line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
 then as much common DNA as possible removed by hybridisation between
 the 2 libraries. Residual EI-specific DNA was amplified, inserted
 into lambda gt10 and screened by plaque hybridisation. A minbank of
 273 E-specific clones was recovered, their inserts amplified and
 used to screen total RNA from both cell types. One clone specific
 for EI was identified and sequenced. The insert from this clone was
 used to screen cDNA prep. from EI RNA and the longest posn. insert
 cloned in pGEM 112f (+) to give PKO164. This insert was sequenced;
 it encoded the 796 AA mouse precursor protein (Q44391/R49730). The
 CC insert was also used to screen a cDNA bank prep. from human
 CC osteosarcoma to identify 2 clones encoding the 2 human precursor
 CC proteins - OSF-4-1 and OSF-4-2 (Q44392/R49731 and Q44393/R49732
 CC respectively).
 SO Sequence 796 AA:

Query Match 54.3%; Score 57; DB 1; Length 796;
 Best Local Similarity 50.0%; Pred. No. 5.92e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 100 IFVIDKSGN1H 111
 |||:|::|:|:
 Oy 3 IFVVDKNTGDIN 14

RESULT 14
 ID W25636 standard: Protein: 796 AA.
 AC W25636:
 DT 03-NOV-1997 (first entry)
 DE Human cadherin-11.
 KW Human; cadherin; rat; calcium-dependent cell adhesion protein;
 KM superfamily; cytoskeleton; eatenin; cancer.
 OS Homo sapiens.
 PN US5646250-A.
 PD 08-JUL-1997.
 PF 17-APR-1992: 872643.
 PR 19-APR-1993: US-049460.
 PR 17-APR-1992: US-872643.
 PR 01-NOV-1994: US-332638.
 PA (DOHE-) DOHENY EYE INST.
 PI Suzuki S;
 DR WPI: 97-362997/33.
 DR N-PSDB: T85403.
 PT Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion
 PS Claim 1: Column 89-94: 56pp: English.
 CC This sequence represents human cadherin-11. The invention specifically
 CC provides details of human cadherin-5, -8, -11, -12 and -13, and rat

CC cadherin-8, -5, -11 and -13. Cadherins are calcium-dependent cell
 CC adhesion proteins. They are glycosylated integral membrane proteins
 CC that have an N-terminal extracellular domain that determines binding
 CC specificity, a hydrophobic membrane spanning region and a C-terminal
 CC cytoplasmic domain, which is highly conserved among members of the
 CC superfamily. The C-terminal domain interacts with the cytoskeleton
 CC through eatens and other cytoskeleton-associated proteins. The
 CC novel cadherin proteins may be used in the analysis of the role of
 CC cadherins in various cancers. Sequence analysis of the cadherin
 CC proteins also allows investigation of the structure and function of
 CC cadherin. The cadherin proteins may be isolated by using anti-cadherin
 CC antibodies. These antibodies may also be used to modulate the activity
 CC of cadherin and to determine the tissue specific distribution of cadherin
 CC proteins. Each subclass of cadherins has a unique tissue distribution
 CC pattern.
 SO Sequence 796 AA:

Query Match 54.3%; Score 57; DB 1; Length 796;
 Best Local Similarity 50.0%; Pred. No. 5.92e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 100 IFVIDKSGN1H 111
 |||:|::|:|:
 Oy 3 IFVVDKNTGDIN 14

RESULT 15
 ID R49730 standard: Protein: 796 AA.
 AC R49730:
 DT 14-SEP-1994 (first entry)
 DE Sequence encoded by murine OSF-4 cDNA.
 KW OSF-4: cadherin; growth factor; osteogenesis; osteoblast; therapy;
 KM diagnosis.
 OS Mus musculus.
 PN EP-585801-A.
 PD 09-MAR-1994.
 PF 25-AUG-1993: 113602.
 PR 28-AUG-1992: JP-230028.
 PA (FARH) HOECHST JAPAN LTD.
 PI Amann E, Kawai S, Okazaki M, Takeshita S, Tsujimura A;
 DR WPI: 94-076152/10.
 DR N-PSDB: 044391/
 PT New bone related, cadherin-like OSF-4 proteins - for treatment
 and diagnosis of bone metabolic disease, and nucleic acid
 encoding them
 PS Claim 1, Page 13-17: 34pp: English.
 CC cDNA libraries were constructed from the mouse osteoblastic cell
 CC line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
 then as much common DNA as possible removed by hybridisation between
 CC the 2 libraries. Residual EI-specific DNA was amplified, inserted
 CC into lambda gt10 and screened by plaque hybridisation. A minbank of
 CC 273 E-specific clones was recovered, their inserts amplified and
 CC used to screen total RNA from both cell types. One clone specific
 CC for EI was identified and sequenced. The insert from this clone was
 CC cloned in pGEM 112f (+) to give PKO164. This insert was sequenced;
 CC it encoded the 796 AA mouse precursor protein (Q44391/R49730). The
 CC insert was also used to screen a cDNA bank prep. from human
 CC osteosarcoma to identify 2 clones encoding the 2 human precursor
 CC proteins - OSF-4-1 and OSF-4-2 (Q44392/R49731 and Q44393/R49732
 CC respectively).
 SO Sequence 796 AA:

Query Match 54.3%; Score 57; DB 1; Length 796;
 Best Local Similarity 50.0%; Pred. No. 5.92e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 100 IFVIDKSGN1H 111
 |||:|::|:|:
 Oy 3 IFVVDKNTGDIN 14

Search completed: Sat May 13 07:05:22 2000

Mon May 15 08:08:02 2000

US-08-991-628-2a.rag

Page 5

Job time : 7 secs.

Journal of Management Education 30(6)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

MSEARCH protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:04:49 2000; Maspar time 4.29 Seconds
Tabular output not generated. 164.911 Million cell updates/sec

Title: >US-08-991-628-2
Description: (1-15) from US08991628.pep
Perfect Score: 105
Sequence: 1 FGIFVVDKNTGDINI 15
Scoring table: PAM 150
Gap 15
Searched: 142080 seqs, 47172406 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 26.968; Variance 38.706; scale 0.697
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description | Pred. No. |
|------------|-------|-------|--------|----|-------------|-----------------------------------|
| 1 | 105 | 100.0 | 999 | 1 | IUHUG3 | desmoglein 3 precursor 1.33e-09 |
| 2 | 84 | 80.0 | 1043 | 1 | IUBOG1 | desmoglein 1 precursor 6.14e-05 |
| 3 | 84 | 80.0 | 1049 | 1 | IUHUG1 | desmoglein 1 precursor 6.14e-05 |
| 4 | 82 | 78.1 | 1117 | 2 | S38673 | desmoglein 2 - human 1.63e-04 |
| 5 | 66 | 62.9 | 761 | 1 | IUBODE | desmoglein 1a - bovl 2.64e-01 |
| 6 | 66 | 62.9 | 785 | 2 | I50180 | cadherin-7 - chicken 2.64e-01 |
| 7 | 66 | 62.9 | 839 | 1 | IJBODF | desmocollin 1b precursor 2.64e-01 |
| 8 | 65 | 61.9 | 770 | 2 | I48910 | desmocollin 1a precursor 4.08e-01 |
| 9 | 65 | 61.9 | 824 | 2 | A48910 | desmocollin 1a precursor 4.08e-01 |
| 10 | 65 | 61.9 | 840 | 2 | I37281 | Desclb precursor - hum 4.08e-01 |
| 11 | 65 | 61.9 | 894 | 2 | I37282 | Desclb precursor - hum 4.08e-01 |
| 12 | 61 | 58.1 | 1612 | 2 | S59869 | DNA topoisomerase (AT 2.22e+00 |
| 13 | 61 | 58.1 | 1626 | 2 | A39242 | DNA topoisomerase (AT 2.22e+00 |
| 14 | 60 | 57.1 | 94 | 2 | A65141 | hypothetical 10.6 KD 3.36e+00 |
| 15 | 60 | 57.1 | 214 | 2 | S76379 | hypothetical protein 3.36e+00 |
| 16 | 60 | 57.1 | 598 | 2 | D69292 | aldehyde ferredoxin o 3.36e+00 |
| 17 | 60 | 57.1 | 790 | 2 | G02676 | cadherin-14 - human 3.36e+00 |
| 18 | 59 | 56.2 | 154 | 1 | B64467 | conserved hypothetical 5.06e+00 |
| 19 | 59 | 56.2 | 245 | 2 | F64465 | hypothetical protein 5.06e+00 |
| 20 | 59 | 56.2 | 452 | 2 | A75213 | tidd protein PAB0154 5.06e+00 |
| 21 | 59 | 56.2 | 624 | 2 | T00044 | vacuolar sorting rece 7.59e+00 |
| 22 | 58 | 55.2 | 270 | 2 | G69469 | conserved hypothetical 7.59e+00 |
| 23 | 58 | 55.2 | 593 | 2 | T04446 | hypothetical protein 7.59e+00 |

| 24 | 58 | 55.2 | 1069 | 2 | T00040 | BH-protocadherin PCDH 7.59e-00 |
|----|----|------|------|---|--------|--------------------------------|
| 25 | 58 | 55.2 | 1072 | 2 | T00041 | BH-protocadherin PCDH 7.59e-00 |
| 26 | 58 | 55.2 | 1200 | 2 | T00042 | BH-protocadherin PCDH 7.59e-00 |
| 27 | 58 | 55.2 | 3027 | 2 | J01917 | polyprotein - parsnip 1.13e+01 |
| 28 | 57 | 54.3 | 209 | 2 | C64685 | ribonuclease H1 - He 1.13e+01 |
| 29 | 57 | 54.3 | 209 | 2 | F71830 | ribonuclease h1 - He 1.13e+01 |
| 30 | 57 | 54.3 | 504 | 1 | G71248 | tidd homolog PH0246 - 1.13e+01 |
| 31 | 57 | 54.3 | 796 | 2 | I48277 | cadherin-11 - mouse 1.13e+01 |
| 32 | 57 | 54.3 | 796 | 2 | I48556 | cadherin-11 - mouse 1.13e+01 |
| 33 | 57 | 54.3 | 796 | 2 | A53584 | OB-cadherin precursor 1.13e+01 |
| 34 | 57 | 54.3 | 796 | 2 | A53992 | cadherin 11 precursor 1.13e+01 |
| 35 | 57 | 54.3 | 798 | 2 | S62791 | probable lipoprotein 1.13e+01 |
| 36 | 57 | 54.3 | 887 | 1 | IJCACL | E-cadherin precursor 1.13e+01 |
| 37 | 57 | 54.3 | 1400 | 1 | I38185 | protein-tyrosine kina 1.13e+01 |
| 38 | 56 | 53.3 | 423 | 2 | T06774 | cell division protein 1.68e+01 |
| 39 | 56 | 53.3 | 713 | 2 | B38992 | cadherin 13 precursor 1.68e+01 |
| 40 | 56 | 53.3 | 756 | 2 | S67433 | hypothetical protein 1.68e+01 |
| 41 | 55 | 52.4 | 211 | 2 | F69391 | hypothetical protein 2.48e+01 |
| 42 | 55 | 52.4 | 591 | 2 | H72474 | probable acylamino-ac 2.48e+01 |
| 43 | 55 | 52.4 | 882 | 1 | H72474 | cadherin 1 precursor 2.48e+01 |
| 44 | 55 | 52.4 | 1196 | 2 | S46430 | botulinum neurotoxin- 2.48e+01 |
| 45 | 55 | 52.4 | 1196 | 2 | J01467 | toxin, nontoxic compo 2.48e+01 |

ALIGNMENTS

RESULT 1
ENTRY IUHUG3 #type complete
TITLE desmoglein 3 precursor - human
ALTERNATE_NAMES pemphigus vulgaris antigen
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS
REFERENCE A41088
#authors Amagai, M.; Klaus-Kovtun, V.; Stanley, J. R.
#journal Cell (1991) 67:869-877
#title Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion.

#cross-references M01D:92069753
#accession A41088
#molecule-type mRNA
#residues 1-999 #label AMA
#cross-references GB:M76482; NRD:9190751; PIDN:AAA60230.1; PID:9190752

GENETICS

#gene GDB:DSG3
#cross-references GDB:134030; OMIM:169615
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE
1-23 #domain signal sequence #status predicted #label SIG
24-49 #domain propeptide #status predicted #label PRO
50-999 #product desmoglein homolog #status predicted #label MAT
50-615 #domain extracellular #status predicted #label EXT
52-157 #domain cadherin repeat homology #label CR1
160-267 #domain cadherin repeat homology #label CR2
270-383 #domain cadherin repeat homology #label CR3
390-495 #domain cadherin repeat homology #label CR4
496-598 #domain cadherin repeat homology #label CR5
616-639 #domain transmembrane #status predicted #label TMN
640-699 #domain intracellular #status predicted #label INT
910-938 #domain desmoglein repeat #label DGL
937-966 #domain desmoglein repeat #label DGL
110,180,545 #binding_site carbohydrate (asn) (covalent) #status predicted

SUMMARY #length 999 #molecular-weight 107502 #checksum 8311

Query Match 100.0%; Score 105; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 1.33e-09;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 97 FGIFVVDKNTGDINI 111
 OY 1 FGIFVVDKNTGDINI 15

RESULT 2
 ENTRY 1JB0G1 #type complete
 TITLE desmoglein 1 precursor - bovine
 ALTERNATE_NAMES desmoglein BDGM
 ORGANISM #formal_name Bos primigenius taurus #common_name cattle
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS S14603: A38872; A37785; S38721; A48173; S24412
 S14603

REFERENCE
 #authors Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #submision submitted to the EMBL Data Library, March 1991
 #description Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.

#accession S14603
 ##molecule_type mRNA
 ##residues 1-1043 #label KOC
 #cross-references EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307

REFERENCE A38872
 #authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1991) 55:200-208
 #title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.

#cross-references EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307
 #accession A38872

#molecule_type mRNA
 ##residues 1-87:968-1043 #label K02
 #cross-references GB:S64268; GB:S64270

REFERENCE A37785
 #authors Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowlin, P.
 #journal Biochem. Biophys. Res. Commun. (1990) 173:1224-1230
 #title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.

#cross-references EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307
 #accession A37785

#molecule_type mRNA
 ##residues 44-123, 'V', 125-493 #label G00
 #cross-references GB:X58165; NID:g162366; PIDN:AAA62709.1; PID:g552318

REFERENCE S38721
 #authors Zimbelmann, R.
 #submision submitted to the EMBL Data Library, February 1991
 #accession S38721

#molecule_type mRNA
 ##residues 44-1043 #label ZIM
 #cross-references EMBL:X57784; NID:g436061; PIDN:CAA40930.1; PID:g436062

REFERENCE A48173
 #authors Koch, P.J.; Walsh, M.J.; Schmeltz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1990) 53:1-12
 #title Identification of desmoglein, a constitutive desmosomal glycoprotein, as a member of the cadherin family of cell adhesion molecules.

#cross-references EMBL:X57784; NID:g436061; PIDN:CAA40930.1; PID:g436062
 #accession A48173

#molecule_type mRNA
 ##residues 44-1001, 'AQPSPAR' #label K03
 #cross-references GB:X57784
 #note this sequence has been revised in references A38872 and S38721

GENETICS DSG1
 #gene Superfamily cadherin: cadherin repeat homology
 CLASSIFICATION calcium binding: cell adhesion: duplication: glycoprotein:
 KEYWORDS

FEATURE 1-23
 24-49
 50-1043
 50-548
 52-157
 160-269
 272-385
 392-491
 549-574
 575-1043
 846-875
 876-905
 906-933
 934-962
 963-1012
 110
 180, 496

transmembrane protein

SUMMARY #length 1043 #molecular_weight 112242 #checksum 6897
 Query Match 80.0%; Score 84; DB 1; Length 1043;
 Best Local Similarity 60.0%; Pred. No. 6,14e-05;
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 97 FGIFVVDKNTGDINI 111
 OY 1 FGIFVVDKNTGDINI 15

RESULT 3
 ENTRY 1JHUG1 #type complete
 TITLE desmoglein 1 precursor - human
 ALTERNATE_NAMES desmosomal glycoprotein I
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS S16906; A39706; A61254; A61279; S16158
 S16906

REFERENCE
 #authors Buxton, R.S.
 #submision submitted to the EMBL Data Library, November 1990
 #accession S16906

#molecule_type mRNA
 ##residues 1-1049 #label BUX
 #cross-references EMBL:X56554; NID:g30505; PIDN:CAA39976.1; PID:g30506

REFERENCE A39706
 #authors Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arneemann, J.; Rutman, A.J.; Pidsley, S.C.; Watt, F.M.; Rees, D.A.; Buxton, R.S.; Magee, A.I.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4796-4800
 #title Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.

#cross-references EMBL:X56554; NID:g30505; PIDN:CAA39976.1; PID:g30506
 #accession A39706

#molecule_type mRNA
 ##residues 24-1049 #label WHE
 #cross-references GB:X56554

REFERENCE A61254
 #authors Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.
 #journal J. Cell Sci. (1991) 99:809-821
 #title Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.

#cross-references EMBL:X56554; NID:g30505; PIDN:CAA39976.1; PID:g30506
 #accession A61254

#molecule_type mRNA
 ##residues 26-1049 #label NIT
 #cross-references GB:X56554; NID:g30505; PIDN:CAA39976.1; PID:g30506

REFERENCE A61279
 #authors Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arneemann, J.; Rees, D.A.; King, I.A.; Magee, A.I.

#journal Biochem. Soc. Trans. (1991) 19:1060-1064
#title Desmosomal glycoproteins I, II and III: novel members of the
cadherin superfamily.
#cross-references MUID:92175187
#accession A61279
#status not compared with conceptual translation
#molecule-type mRNA
#residues 1-55 ##label WH3

GENETICS
#gene GDB:DSG1
#cross-references GDB:126563; OMIM:125670
#map-position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein

FEATURE
1-23
24-49 #domain signal sequence #status predicted #label SIG
50-1049 #domain propeptide #status predicted #label PRO
50-548 #product desmoglein #status predicted #label MAT
52-157 #domain extracellular #status predicted #label EXT
160-269 #domain cadherin repeat homology #label CR1
272-385 #domain cadherin repeat homology #label CR2
392-493 #domain cadherin repeat homology #label CR3
509-530 #region serine/threonine-rich
549-569 #domain transmembrane #status predicted #label TM
572-1049 #domain intracellular #status predicted #label INT
840-869 #domain desmoglein repeat #label DG1
870-899 #domain desmoglein repeat #label DG2
900-927 #domain desmoglein repeat #label DG3
928-956 #domain desmoglein repeat #label DG4
969-1019 #region glycine/serine-rich
110-180 #binding-site carbohydrate (asn) (covalent) #status
predicted

SUMMARY #length 1049 #molecular-weight 113715 #checksum 4482

Query Match 80.0%; Score 84; DB 1; Length 1049;
Best Local Similarity 60.0%; Pred. No. 6; 14e-05;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 YGIFVFNKDTGELNY 111
Oy 1 FGIFVFNKDTGELNY 15

RESULT 4
ENTRY S38673 #type complete
TITLE desmoglein 2 - human
ALTERNATE_NAMES #formal_name Homo sapiens #common_name man
ORGANISM 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
DATE 20-Aug-1999
ACCESSIONS S38673; B38872
REFERENCE S38673
#authors Zimbelmann, R.
#submission submitted to the EMBL Data Library, September 1993
#accession S38673
#status preliminary
#molecule-type mRNA
#residues 1-1117 ##label ZIM
#cross-references EMBL:Z26317; NID:g416177; PIDN:CAA81226.1;
#cross-references MUID:9416178

REFERENCE A38872
#authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.;
#journal Eur. J. Cell Biol. (1991) 55:200-208
#title Complete amino acid sequence of the epidermal desmoglein
precursor polypeptide and identification of a second type
of desmoglein gene.
#cross-references MUID:92037656
#accession B38872
#molecule-type mRNA
#residues 777-1117 ##label KOC

#cross-references GB:S64273

GENETICS
#gene GDB:DSG2
#cross-references GDB:128808; OMIM:125671
#map-position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
membrane protein

FEATURE
51-158 #domain cadherin repeat homology #label CR1
161-271 #domain cadherin repeat homology #label CR2
SUMMARY #length 1117 #molecular-weight 122384 #checksum 7660

Query Match 78.1%; Score 82; DB 2; Length 1117;
Best Local Similarity 60.0%; Pred. No. 1; 63e-04;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 98 YGIFVFNKDTGELNY 112
Oy 1 FGIFVFNKDTGELNY 15

RESULT 5
ENTRY IJBODE #type complete
TITLE desmocollin 1a - bovine
ALTERNATE_NAMES desmocollin BOCM; desmosomal glycoprotein 2
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
22-Jun-1999
ACCESSIONS A43838; A38456; A60714; S14542
REFERENCE A43838
#authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.;
Schmeitz, M.; Franke, W.W.
#journal Differentiation (1991) 47:29-36
#title Amino acid sequence of bovine muzzle epithelial desmocollin
derived from cloned cDNA: a novel subtype of desmosomal
cadherins.
#cross-references MUID:92008912
#accession A43838
#molecule-type mRNA
#residues 1-761 ##label KOC
#cross-references GB:X58029; GB:S57985; NID:g453583; PIDN:CAA1088.1;
#cross-references MUID:9453584
#accession B43838
#molecule-type protein
#residues 1-32; 65-76; 148-159; 164-176; 190-205; 208-219; 238-256;
#residues 361-375; 377-388; 478-486 ##label KOC
#experimental_source muzzle epithelium
#note sequence extracted from NCBI backbone
REFERENCE A38456
#authors Collins, J.E.; Legan, P.K.; Kenny, T.P.; Macgarvie, J.;
Hollton, J.L.; Garrard, D.R.
#journal J. Cell Biol. (1991) 113:381-391
#title Cloning and sequence analysis of desmosomal glycoproteins 2
and 3 (desmocollins): cadherin-like desmosomal adhesion
molecules with heterogeneous cytoplasmic domains.
#cross-references MUID:91185414
#accession A38456
#molecule-type mRNA
#residues 606-761 ##label COL
#cross-references EMBL:X56967; NID:g310; PIDN:CAA40287.1; PID:g311
#accession A60714
REFERENCE A60714
#authors Hollton, J.L.; Kenny, T.P.; Legan, P.K.; Collins, J.E.; Keen,
J.N.; Sharma, R.; Garrard, D.R.
#journal J. Cell Sci. (1990) 97:239-246
#title Desmosomal glycoproteins 2 and 3 (desmocollins) show
N-terminal similarity to calcium-dependent cell-cell
adhesion molecules.
#cross-references MUID:9115997
#accession A60714
#molecule-type protein
#residues 1-6; A', 8-9, 'R', 11-17, 'RCE', 21-23 ##label HOL
#experimental_source nasal epidermis

```

GENETICS
#gene
CLASSIFICATION
KEYWORDS
DSCL
#superfamily cadherin; cadherin repeat homology
#alternative splicing; calcium binding; cell adhesion;
#duplication; glycoprotein; phosphoprotein; transmembrane
#protein
FEATURE
1-761
1-561
3-108
111-120
223-338
339-444
445-561
563-582
583-761
31,266,413
584,588,678
605
671
681
682
SUMMARY
#length 761 #molecular-weight 85170 #checksum 3508
Query Match 62.94; Score 66; DB 1; Length 761;
Best Local Similarity 53.88; Pred. No. 2,646-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Db 48 FNLFIEKDTGDI 60
1 : : : : :
QY 1 GFIFVDKNTGDI 13
RESULT 6
ENTRY 150180 #type complete
TITLE cadherin-7 - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change
20-Aug-1999
ACCESSIONS
REFERENCE
150180
150178
#authors Nakagawa, S.; Takeichi, M.
#journal Development (1995) 121:1321-1332
#title Neural crest cell-cell adhesion controlled by sequential and
#subpopulation-specific expression of novel cadherins.
#cross-references M01D:95309115
#accession 150180
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type MRNA
#residues 1-785 #label NAK
#cross-references GB:D42350; NID:g8688000; PID:BA07721.1; PID:g868001
CLASSIFICATION
SUMMARY
#superfamily cadherin; cadherin repeat homology
#length 785 #molecular-weight 87171 #checksum 8490
Query Match 62.94; Score 66; DB 2; Length 785;
Best Local Similarity 61.54; Pred. No. 2,646-01;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 93 SFIIDENIGDTH 105
: : : : :
QY 2 GFIFVDKNTGDI 14
RESULT 7
ENTRY IJBODF #type complete
TITLE desmocollin 1b precursor - bovine
#alternate_names desmosomal glycoprotein 3

```

```

ORGANISM
DATE
#formal_name Bos primigenius laurus #common_name cattle
30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
22-Jun-1999

ACCESSIONS
B38456: A39377: S14567

REFERENCE
#authors
Collins, J.E.; Legan, P.K.; Kenny, T.P.; MacGarvie, J.;
Holton, J.L.; Garrod, D.R.
J. Cell Biol. (1991) 113:381-391
Cloning and sequence analysis of desmosomal glycoproteins 2
and 3 (desmocollins): cadherin-like desmosomal adhesion
molecules with heterogeneous cytoplasmic domains.

#cross-references MUID:91185414
#accession B38456
#molecule_type mRNA
#residues 1-839 #label COL
#cross-references GB:X56966; MID:g315; PIDN:CAA40286.1; PID:g316
A39377
REFERENCE
#authors
Mechanic, S.; Raynor, K.; Hill, J.E.; Cowlin, P.
Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4476-4480
Desmocollins form a distinct subset of the cadherin family of
cell adhesion molecules.
#cross-references MUID:91239591
#accession A39377
#molecule_type mRNA
#residues 1-484, 'A', 486-839 #label MEC
#cross-references GB:M67489; GB:M61750; MID:g162970; PIDN:AAA30492.1;
PID:g162971
part of this sequence, including the amino end of the
mature protein, was confirmed by protein sequencing

GENETICS
#note
part of this sequence, including the amino end of the
mature protein, was confirmed by protein sequencing

CLASSIFICATION
#gene
DSC1
#keywords
'superfamily cadherin; cadherin repeat homology
alternative splicing; calcium binding; cell adhesion;
duplication; glycoprotein; phosphoprotein; transmembrane
protein

FEATURE
1-29
30-132 #domain signal sequence #status predicted #label SIC\
133-839 #domain propeptide #status predicted #label PRO\
133-889 #product desmocollin 1b #status experimental #label XTC\
133-682 #domain extracellular #status predicted #label EXT\
135-240 #domain cadherin repeat homology #label CR1\
243-352 #domain cadherin repeat homology #label CR2\
355-470 #domain cadherin repeat homology #label CR3\
471-576 #domain cadherin repeat homology #label CR4\
577-682 #domain cadherin repeat homology #label CR5\
684-714 #domain transmembrane #status predicted #label TM\
718-839 #domain intracellular #status predicted #label INT\
163,398,545 #binding_site carbohydrate (Asn) (covalent) #status
predicted\
716,720,810 #binding_site phosphate (Thr) (covalent) (by protein
kinase C) #status predicted\
737 #binding_site phosphate (Ser) (covalent) (by casein
kinase II) #status predicted\
803,830 #binding_site phosphate (Ser) (covalent) (by protein
kinase C) #status predicted\
813 #binding_site phosphate (Tyr) (covalent) #status
predicted\
814 #binding_site phosphate (Thr) (covalent) (by casein
kinase II) #status predicted
SUMMARY
#length 839 #molecular-weight 93521 #checksum 6365

Query Match 62.9%; Score 66; DB 1; Length 839;
Best Local Similarity 53.8%; Pred. No. 2,64e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 180 FNLFIEKDTGDI 192
1 : 1 : 1 : 1 : 1 : 1 :
QY 1 FGIFVVDKNTGDI 13

RESULT 8
ENTRY B48910 #type fragment
TITLE desmocollin 1b precursor - human (fragment)

```

ORGANISM #formal_name Homo sapiens #common_name man
DATE 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change
13-Aug-1997

ACCESSIONS B48910
REFERENCE #authors King, I.A.; Arnemann, J.; Spurr, N.K.; Buxton, R.S.
#journal Genomics (1993) 18:185-194
#title Cloning of the cDNA (DSCI) coding for human type 1 desmocollin and its assignment to chromosome 18.

#cross-references MUID:94116981
#accession B48910
#status Preliminary
#molecule_type mRNA
#residues 1-770 #label KIN
#cross-references GB:X72929

GENETICS
#gene GDB:DSCI
#cross-references GDB:128632
#map_position 18q12.2-18q12.2
CLASSIFICATION #superfamily_cadherin: cadherin repeat homology
KEYWORDS #alternative_splicing: calcium binding; cell adhesion; glycoprotein; phosphoprotein; transmembrane protein

FEATURE 67-172 #domain_cadherin repeat homology #label CR1
SUMMARY #length 770 #checksum 4861

Query Match 61.9% Score 65; DB 2; Length 770;
Best Local Similarity 53.8%; Pred. No. 4.08e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 112 FNLFYIKDGTGI 124
1 : : : : :
QY 1 FGIFVVDKNTGDI 13

RESULT 9
ENTRY A48910 #type fragment
TITLE desmocollin 1a precursor - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change
15-Aug-1997

ACCESSIONS A48910
REFERENCE #authors King, I.A.; Arnemann, J.; Spurr, N.K.; Buxton, R.S.
#journal Genomics (1993) 18:185-194
#title Cloning of the cDNA (DSCI) coding for human type 1 desmocollin and its assignment to chromosome 18.

#cross-references MUID:94116981
#accession A48910
#status Preliminary
#molecule_type mRNA
#residues 1-824 #label KIN
#cross-references GB:X72925

GENETICS
#gene GDB:DSCI
#cross-references GDB:128632
#map_position 18q12.2-18q12.2
CLASSIFICATION #superfamily_cadherin: cadherin repeat homology
KEYWORDS #alternative_splicing: calcium binding; cell adhesion; glycoprotein; phosphoprotein; transmembrane protein

FEATURE 67-172 #domain_cadherin repeat homology #label CR1
SUMMARY #length 824 #checksum 2645

Query Match 61.9% Score 65; DB 2; Length 824;
Best Local Similarity 53.8%; Pred. No. 4.08e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 112 FNLFYIKDGTGI 124
1 : : : : :
QY 1 FGIFVVDKNTGDI 13

RESULT 10
ENTRY I37281 #type complete
TITLE Dscia precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
20-Aug-1999

ACCESSIONS I37281
REFERENCE #authors Theis, D.G.; Koch, P.J.; Franke, W.W.
#journal Int. J. Dev. Biol. (1993) 37:101-110
#title Differential synthesis of type 1 and type 2 desmocollin mRNAs in human stratified epithelia.

#cross-references MUID:93283249
#accession I37281
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-840 #label RES
#cross-references EMBL:234532; NID:g505536; PIDN:CAA84278.1; PID:g505537

CLASSIFICATION #superfamily_cadherin: cadherin repeat homology
SUMMARY #length 840 #molecular_weight 93848 #checksum 3174

Query Match 61.9% Score 65; DB 2; Length 840;
Best Local Similarity 53.8%; Pred. No. 4.08e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 182 FNLFYIKDGTGI 194
1 : : : : :
QY 1 FGIFVVDKNTGDI 13

RESULT 11
ENTRY I37282 #type complete
TITLE Dscib precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
20-Aug-1999

ACCESSIONS I37282
REFERENCE #authors Theis, D.G.; Koch, P.J.; Franke, W.W.
#journal Int. J. Dev. Biol. (1993) 37:101-110
#title Differential synthesis of type 1 and type 2 desmocollin mRNAs in human stratified epithelia.

#cross-references MUID:93283249
#accession I37282
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-894 #label RES
#cross-references EMBL:234532; NID:g505536; PIDN:CAA84279.1; PID:g505538

GENETICS
#introns 829/2
CLASSIFICATION #superfamily_cadherin: cadherin repeat homology
SUMMARY #length 894 #molecular_weight 100044 #checksum 7946

Query Match 61.9% Score 65; DB 2; Length 894;
Best Local Similarity 53.8%; Pred. No. 4.08e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 182 FNLFYIKDGTGI 194
1 : : : : :
QY 1 FGIFVVDKNTGDI 13

RESULT 12
ENTRY S5969 #type complete
TITLE DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) isoform beta - Chinese hamster
ALTERNATE_NAMES DNA topoisomerase II isoform beta; DNA-gyrase
ORGANISM #formal_name Crictetus griseus #common_name Chinese hamster
DATE 15-Feb-1996 #sequence_revision 13-Mar-1997 #text_change
16-Jul-1999

ACCESSIONS S5969; S54154

REFERENCE S59966
#authors Derudder, S.; Frey, S.; Delaporte, C.; Jacquemin-Sablon, A.
#journal Blochim. Biophys. Acta (1995) 1264:178-182
#title Cloning and characterization of full-length cDNAs coding for the DNA topoisomerase II beta from Chinese hamster lung cells sensitive and resistant to 9-OH-ellipticine.
#cross-references EMBL:96085121
#accession S59966
#status translation not shown
#molecule-type mRNA
#residues 1-1612 ##label DER
#cross-references EMBL:X86455; NID:9790987; PIDN:CAA60173.1; PID:9790988

##experimental_source lung
CLASSIFICATION #superfamily eukaryotic type II DNA topoisomerase: phage T4
DNA topoisomerase (ATP-hydrolyzing) medium chain homology
KEYWORDS ATP; DNA binding; DNA replication; isomerase; nucleus
FEATURE
697-927 #domain phage T4 DNA topoisomerase (ATP-hydrolyzing)
medium chain homology #label T4T
#length 1612 #molecular-weight 182074 #checksum 5332

SUMMARY
Query Match 58.1%; Score 61; DB 2; Length 1612;
Best Local Similarity 72.7%; Pred. No. 2,22e+00;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 922 IFVVDRTVEI 932
|||||:|:
3 IFVVDKNTGDI 13

RESULT 13
ENTRY A39242 #type complete
TITLE DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) beta,
splice form 2 - human
ALTERNATE_NAMES DNA topoisomerase II isoform beta-2
CONTAINS DNA topoisomerase II isoform beta-1
ORGANISM #formal name Homo sapiens #common name man
DATE 04-Oct-1991 #sequence-revision 03-May-1996 #text-change 16-Jul-1999

ACCESSIONS S26730; A39242; S10710; S33970; S30191; S41641; S30190
#authors Jenkins, J.R.; Aylon, P.; Jones, T.; Davies, S.L.; Simmons, D.L.; Harris, A.L.; Sheer, D.; Hickson, I.D.
#journal Nucleic Acids Res. (1992) 20:5587-5592
#title Isolation of cDNA clones encoding the beta isozyme of human DNA topoisomerase II and localisation of the gene to chromosome 3p24.
#cross-references MUID:93087165
#accession S26730
#molecule-type mRNA
#residues 1-23,29-1626 ##label JEN
#cross-references EMBL:X68060; NID:937230; PIDN:CAA48197.1; PID:937231

REFERENCE A39242
#authors Chung, T.D.Y.; Drake, F.H.; Tan, K.B.; Per, S.R.; Crooke, S.T.; Mirabeli, C.K.
#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:9431-9435
#title Characterization and Immunological identification of cDNA clones encoding two human DNA topoisomerase II isozymes.
#cross-references MUID:90083281
#accession A39242
#molecule-type mRNA
#residues 149-1043 ##label CHU
#cross-references GB:M27504
#journal S10710
#authors Austin, C.A.; Fisher, L.M.
#journal FEBS Lett. (1990) 266:115-117
#title Isolation and characterization of a human cDNA clone encoding a novel DNA topoisomerase II homologue from HeLa cells.
#cross-references MUID:90306333
#accession S10710
#molecule-type mRNA
#residues 1043-1276 ##label AUS

##cross-references GB:X53662; GB:S56813; NID:938324; PIDN:CAA37706.1; PID:938325

REFERENCE S30190
#authors Austin, C.A.; Sng, J.H.; Patel, S.; Fisher, L.M.
#journal Blochim. Biophys. Acta (1993) 1172:283-291
#title Novel HeLa topoisomerase II is the II-beta isoform: complete coding sequence and homology with other type II topoisomerases.
#cross-references MUID:93192319
#accession S33970
#molecule-type mRNA
#residues 1-23,29-1610,'A',1612-1626 ##label A02
#cross-references EMBL:215111

REFERENCE S30191
#authors Austin, C.A.; Sng, J.H.; Patel, S.; Fisher, L.M.
#submission submitted to the EMBL Data Library, September 1992
#accession S30191
#molecule-type mRNA
#residues 596-1430,'S',1432-1610,'A',1612-1626 ##label A01
#cross-references EMBL:215115; NID:9288564; PIDN:CAA78821.1; PID:9288565

REFERENCE S41641
#authors Davies, S.L.; Jenkins, J.R.; Hickson, I.D.
#journal Nucleic Acids Res. (1993) 21:3719-3723
#title Human cells express two differentially spliced forms of topoisomerase II-beta mRNA.
#cross-references MUID:93376494
#accession S41641
#molecule-type DNA
#residues 24-80 ##label DAV
#cross-references EMBL:X71911
#note this sequence represents a long minor splice form, designated beta-2

GENETICS
#gene GDB:TOP2B
#cross-references GDB:131575; OMIM:126431
#map_position 3p24-3p24
CLASSIFICATION #superfamily eukaryotic type II DNA topoisomerase: phage T4
DNA topoisomerase (ATP-hydrolyzing) medium chain homology
KEYWORDS alternative splicing; ATP; dimer; isomerase; nucleus
FEATURE
1-1626
#product DNA topoisomerase II beta-2 #status predicted
#label MINR
#product DNA topoisomerase II beta-1 #status predicted
#label MAJR
#domain phage T4 DNA topoisomerase (ATP-hydrolyzing)
medium chain homology #label T4T
#length 1626 #molecular-weight 183296 #checksum 3565

SUMMARY
Query Match 58.1%; Score 61; DB 2; Length 1626;
Best Local Similarity 72.7%; Pred. No. 2,22e+00;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 934 IFVVDRTVEI 944
|||||:|:
3 IFVVDKNTGDI 13

RESULT 14
ENTRY A65141 #type complete
TITLE hypothetical 10.6 kb protein in gntR-gvt intergenic region -
Escherichia coli (strain K-12)
ORGANISM #formal name Escherichia coli
DATE 12-Sep-1997 #sequence-revision 17-Sep-1997 #text-change 14-Nov-1997

ACCESSIONS A65141
#journal A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462

```

#title      The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97425617
#accession  A65141
#status     preliminary; nucleic acid sequence not shown;
            translation not shown
##molecule_type DNA
##residues  1-94 #label BLAT
##cross-references GB:AE000421; GB:U00096; NID:g1789854; PID:g1789855;
            UWGP:D3446
##experimental_source strain K-12, substrain MG1655
GENETICS
#gene       yrb
SUMMARY     #length 94 #molecular-weight 10613 #checksum 4152
Query Match 57.1%; Score 60; DB 2; Length 94;
Best Local Similarity 45.5%; Pred. No. 3.36e+00;
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
Db 59 FIIDSGEIH 69
|:|:|:|:|:|:|
OY 4 FVVDKNTGIDIN 14

RESULT 15
ENTRY     S76379 #type complete
TITLE     hypothetical protein - Synecocystis sp. (strain PCC 6803)
ORGANISM  #formal name Synecocystis sp.
           PCC 6803
DATE      25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
           21-Aug-1998
ACCESSIONS
REFERENCE  S76379
#authors   Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
           Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugitara, E.;
           Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
           Muraki, A.; Nakazaki, N.; Naito, K.; Okumura, S.; Shilpo,
           S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
           Yasuda, M.; Tabata, S.
           DNA Res. (1996) 3:109-136
#journal   Sequence analysis of the genome of the unicellular
#title     cyanobacterium Synecocystis sp. PCC6803. II. Sequence
           determination of the entire genome and assignment of
           potential protein-coding regions.
#cross-references MUID:97061201
#accession S76379
#status     preliminary
##molecule_type DNA
##residues  1-214 #label KAN
##cross-references EMBL:D64000; GB:AB001339; NID:g1001484; PID:d1010882;
            PID:g1001603
#note      the nucleotide sequence was submitted to the EMBL Data
            Library, June 1996
SUMMARY     #length 214 #molecular-weight 24539 #checksum 4859
Query Match 57.1%; Score 60; DB 2; Length 214;
Best Local Similarity 42.9%; Pred. No. 3.36e+00;
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
Db 49 FGVLMDPSTGEIS 62
|:|:|:|:|:|:|
OY 1 FGTFVDKNTGIDIN 14

```

Search completed: Sat May 13 07:04:57 2000
 Job time : 8 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 06:58:15 2000: Maspar time 99.86 Seconds
Tabular output not generated. 4.575 Million cell updates/sec

Title: >US-08-991-628-2
Description: (1-15) from US08991628.pep
Perfect Score: 105
Sequence: 1 FGIFFVDKNTGDIINI 15

Scoring table: PAM 150
Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 27.553; Variance 34.618; scale 0.796

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|-----------------------------------|-----------|
| 1 | 105 | 100.0 | 999 | 1 | DSG3_HUMAN DESMOGLEIN 3 PRECURSOR | 2.32e+11 |
| 2 | 84 | 80.0 | 1043 | 1 | DSG1_BOVIN DESMOGLEIN 1 PRECURSOR | 4.39e+06 |
| 3 | 84 | 80.0 | 1049 | 1 | DSG1_HUMAN DESMOGLEIN 1 PRECURSOR | 4.29e+06 |
| 4 | 82 | 78.1 | 1117 | 1 | DSG2_HUMAN DESMOGLEIN 2 PRECURSOR | 1.28e+05 |
| 5 | 66 | 62.9 | 893 | 1 | DSG1_BOVIN DESMOGLEIN 1A/1B PREC | 5.19e+02 |
| 6 | 65 | 61.9 | 894 | 1 | DSG1_HUMAN DESMOGLEIN 1A/1B PREC | 8.44e+02 |
| 7 | 61 | 58.1 | 886 | 1 | DSG1_MOUSE DESMOGLEIN 1A/1B PREC | 5.62e+01 |
| 8 | 61 | 58.1 | 1612 | 1 | TP2B_MOUSE DNA TOPOISOMERASE II. | 5.62e+01 |
| 9 | 61 | 58.1 | 1612 | 1 | TP2B_CRILLO DNA TOPOISOMERASE II. | 5.62e+01 |
| 10 | 61 | 58.1 | 1626 | 1 | TP2B_HUMAN DNA TOPOISOMERASE II. | 5.62e+01 |
| 11 | 61 | 58.1 | 1627 | 1 | TP2B_CHICK DNA TOPOISOMERASE II. | 5.62e+01 |
| 12 | 60 | 57.1 | 94 | 1 | YRHB_ECOLI HYPOTHEICAL 10.6 KD P | 8.91e+01 |
| 13 | 60 | 57.1 | 790 | 1 | CAD6_HUMAN CADHERIN-6 PRECURSOR. | 2.21e+00 |
| 14 | 58 | 55.2 | 3027 | 1 | POLG_PTYVI GENOME POLYPROTEIN (CO | 3.44e+00 |
| 15 | 57 | 54.3 | 209 | 1 | RNH2_HELPY RIBONUCLEASE HII (EC 3 | 3.44e+00 |
| 16 | 57 | 54.3 | 209 | 1 | RNH2_HELPY RIBONUCLEASE HII (EC 3 | 3.44e+00 |
| 17 | 57 | 54.3 | 796 | 1 | CAD6_MOUSE CADHERIN-6 PRECURSOR | 3.44e+00 |
| 18 | 57 | 54.3 | 796 | 1 | CAD6_HUMAN CADHERIN-6 PRECURSOR | 3.44e+00 |
| 19 | 57 | 54.3 | 887 | 1 | CAD1_CHICK EPITHELIAL-CADHERIN PR | 3.44e+00 |
| 20 | 57 | 54.3 | 1400 | 1 | RON_HUMAN MACROPHAGE-STIMULATING | 3.44e+00 |
| 21 | 56 | 53.3 | 433 | 1 | FTSZ_ARATH CELL DIVISION PROTEIN | 5.34e+00 |
| 22 | 56 | 53.3 | 515 | 1 | PDI_ASPOF PROTEIN DISULFIDE ISOM | 5.34e+00 |
| 23 | 56 | 53.3 | 713 | 1 | CADD_HUMAN T-CADHERIN PRECURSOR (| 5.34e+00 |

| | | | | | | |
|----|----|------|------|---|-----------------------------------|----------|
| 24 | 56 | 53.3 | 756 | 1 | Y034_SCHPO HYPOTHEICAL 81.7 KD P | 5.34e+00 |
| 25 | 56 | 53.3 | 789 | 1 | CADA_CHICK CADHERIN-10 PRECURSOR. | 5.34e+00 |
| 26 | 55 | 52.4 | 566 | 1 | YGA3_YEAST HYPOTHEICAL 62.8 KD T | 8.23e+00 |
| 27 | 55 | 52.4 | 732 | 1 | CAD6_CHICK B-CADHERIN PRECURSOR (| 8.23e+00 |
| 28 | 55 | 52.4 | 789 | 1 | CAD6_RAT CADHERIN-6 PRECURSOR (| 8.23e+00 |
| 29 | 55 | 52.4 | 790 | 1 | CAD6_HUMAN CADHERIN-6 PRECURSOR (| 8.23e+00 |
| 30 | 55 | 52.4 | 829 | 1 | CAD3_HUMAN PLACERIN-6 PRECURSOR (| 8.23e+00 |
| 31 | 55 | 52.4 | 882 | 1 | CAD1_HUMAN EPITHELIAL-CADHERIN PR | 8.23e+00 |
| 32 | 55 | 52.4 | 884 | 1 | CAD1_MOUSE EPITHELIAL-CADHERIN PR | 8.23e+00 |
| 33 | 55 | 52.4 | 965 | 1 | AMPN_MOUSE AMINOPEPTIDASE N (EC 3 | 8.23e+00 |
| 34 | 55 | 52.4 | 1196 | 1 | BXCN_CLOBO BOTULINUM NEUROTOXIN T | 8.23e+00 |
| 35 | 55 | 52.4 | 1615 | 1 | RRPO_TMYTO PUTATIVE RNA-DIRECTED | 8.23e+00 |
| 36 | 55 | 52.4 | 1826 | 1 | SUIS_RABIT SUCRASE-ISOMALTASE, IN | 8.23e+00 |
| 37 | 55 | 52.4 | 1826 | 1 | SUIS_HUMAN SUCRASE-ISOMALTASE, IN | 8.23e+00 |
| 38 | 54 | 51.4 | 185 | 1 | YXAK_BACSU YXAK-19.5 KD P | 1.26e+01 |
| 39 | 54 | 51.4 | 273 | 1 | HPLA_HAEIN 28 KD OUTER MEMBRANE P | 1.26e+01 |
| 40 | 54 | 51.4 | 353 | 1 | YXAL_BACSU HYPOTHEICAL 38.5 KD P | 1.26e+01 |
| 41 | 53 | 50.5 | 305 | 1 | RPO4_VARY DNA-DIRECTED RNA POLYM | 1.92e+01 |
| 42 | 53 | 50.5 | 305 | 1 | RPO4_VACC DNA-DIRECTED RNA POLYM | 1.92e+01 |
| 43 | 53 | 50.5 | 692 | 1 | Y650_METJA HYPOTHEICAL PROTEIN M | 1.92e+01 |
| 44 | 53 | 50.5 | 794 | 1 | CADQ_HUMAN BRAIN-CADHERIN PRECURS | 1.92e+01 |
| 45 | 53 | 50.5 | 2647 | 1 | ABP2_HUMAN ENDOTHELIAL ACTIN-BIND | 1.92e+01 |

ALIGNMENTS

| RESULT ID | DSG3_HUMAN | STANDARD | PRT | 999 AA. |
|-----------|--|----------|-----|---------|
| AC | P32926; | | | |
| DT | 01-OCT-1993 (Rel. 27, Created) | | | |
| DT | 01-OCT-1993 (Rel. 27, Last sequence update) | | | |
| DT | 01-NOV-1997 (Rel. 35, Last annotation update) | | | |
| DE | DESMOGLIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA). | | | |
| GN | DSG3. | | | |
| OS | Homo sapiens (Human). | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; | | | |
| CC | Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE: 92069753. | | | |
| RA | Amagat M., Klaus-Kovtun V., Stanley J.R.; | | | |
| RT | "Antibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion." | | | |
| RL | Cell 67:869-877(1991). | | | |
| CC | -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS. | | | |
| CC | INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE | | | |
| CC | FILAMENTS MEDIATING CELL-CELL ADHESION. | | | |
| CC | -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. | | | |
| CC | -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND | | | |
| CC | CARCINOMAS. | | | |
| CC | -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS | | | |
| CC | (POTENTIAL). | | | |
| CC | -1- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN | | | |
| CC | LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES | | | |
| CC | AGAINST DSG3. | | | |
| CC | -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY. | | | |
| CC | THIS SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| DR | EMBL: M76482; AAA60230.1; - | | | |
| DR | PIR: A41088; ICHUG3. | | | |
| DR | HSP: P09803; IEDH. | | | |
| DR | MIR: 169615. | | | |
| DR | PROSITE: PS00232; CADHERIN: 3. | | | |
| DR | PFAM: PF00028; cadherin: 4. | | | |
| KW | Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein; | | | |

KM Calcium-binding; Repeat.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 49 POTENTIAL.
 FT CHAIN 50 999 DESMOGLEIN 3.
 FT DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 616 640 POTENTIAL.
 FT DOMAIN 641 999 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 50 158 CADHERIN 1.
 FT REPEAT 159 268 CADHERIN 2.
 FT REPEAT 269 383 CADHERIN 3.
 FT REPEAT 386 499 CADHERIN 4.
 FT REPEAT 910 935 DESMOGLEIN REPEAT 1.
 FT REPEAT 936 966 DESMOGLEIN REPEAT 2.
 FT CARBOHYD 110 110 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 459 459 POTENTIAL.
 FT CARBOHYD 545 545 POTENTIAL.
 SQ SEQUENCE 999 AA: 107503 MW: 60479DD46AC219A1 CRC64;

Query Match 100.0%; Score 105; DB 1; Length 999;
 Best Local Similarity 100.0%; Pred. No. 2,32e-11;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 97 FGIFVVDKNTGDIINI 111
 0Y 1 FGIFVVDKNTGDIINI 15

RESULT 2
 ID DSG1_BOVIN STANDARD: PRT: 1043 AA.
 AC 003763;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DSG1).
 OS DSG1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MUSCLE EPITHELIUM;
 RA Koch P.J., Goldschmidt M.D., Zimbelmann R., Franke W.W.;
 RA Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE OF 44-1043 FROM N.A.
 RC TISSUE=MUSCLE EPITHELIUM;
 RA MEDLINE: 9116965.
 RA Koch P.J., Walsh M.J., Schmelz M., Goldschmidt M.D.,
 RA Zimbelmann R., Franke W.W.;
 RT Identification of desmoglein, a constitutive desmosomal
 RT glycoprotein, as a member of the cadherin family of cell adhesion
 RT molecules.*;
 RL Eur. J. Cell Biol. 53:1-12(1990).
 RN [3]
 RP REVISIONS, AND SEQUENCE OF 101-123.
 RA MEDLINE: 92037656.
 RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R.,
 RA Franke W.W.;
 RT Complete amino acid sequence of the epidermal desmoglein precursor
 RT polypeptide and identification of a second type of desmoglein gene.*;
 RL Eur. J. Cell Biol. 55:200-208(1991).
 RN [4]
 RP SEQUENCE OF 44-493 FROM N.A.
 RA MEDLINE: 91097553.
 RA Goodwin L., Hill J.E., Raynor K., Raazi L., Manabe M., Cowin P.;
 RT Desmoglein shows extensive homology to the cadherin family of cell
 RT adhesion molecules.*;
 RA Wheeler G.N., Parker A.E., Thomas C.L., Ataliotis P., Poynter D.,
 RA Blochem. Biophys. Res. Commun. 173:1224-1230(1990).
 CC -I- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -I- TISSUE SPECIFICITY: EPIDERMIS, MIZZLE, TONGUE AND ESOPHAGUS.
 CC -I- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -I- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOsome SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X58466; CA41380.1; -;
 DR EMBL: X57784; CA40930.1; -;
 DR EMBL: M58165; AAA62709.1; -;
 DR PIR: S14603; IJBOG1.
 DR HSSP: P09803; 1EDH.
 DR PROSITE: PS00232; CADHERIN: 2.
 DR PFAM: PF00028; cadherin: 3.
 DR Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
 KM Calcium-binding; Repeat.
 KM SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 49 POTENTIAL.
 FT CHAIN 50 1043 DESMOGLEIN 1.
 FT DOMAIN 50 548 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 549 573 POTENTIAL.
 FT DOMAIN 574 1043 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 50 158 CADHERIN 1.
 FT REPEAT 159 270 CADHERIN 2.
 FT REPEAT 271 385 CADHERIN 3.
 FT REPEAT 386 498 CADHERIN 4.
 FT REPEAT 819 845 DESMOGLEIN REPEAT 1.
 FT REPEAT 846 875 DESMOGLEIN REPEAT 2.
 FT REPEAT 876 905 DESMOGLEIN REPEAT 3.
 FT REPEAT 906 933 DESMOGLEIN REPEAT 4.
 FT REPEAT 934 962 DESMOGLEIN REPEAT 5.
 FT DOMAIN 963 1012 GLY/SER-RICH.
 FT CARBOHYD 110 110 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 496 496 POTENTIAL.
 FT CONFLICT 124 124 I -> V (IN REF. 4).
 SQ SEQUENCE 1043 AA: 112243 MW: ADE46133FB87C11 CRC64;

Query Match 80.0%; Score 84; DB 1; Length 1043;
 Best Local Similarity 60.0%; Pred. No. 4,29e-06;
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 YGIFVVDKNTGDIINI 111
 0Y 1 FGIFVVDKNTGDIINI 15

RESULT 3
 ID DSG1_HUMAN STANDARD: PRT: 1049 AA.
 AC 002413;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DSG1).
 OS DSG1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KERATINOCYTES;
 RA MEDLINE: 91271279.
 RA Wheeler G.N., Parker A.E., Thomas C.L., Ataliotis P., Poynter D.,
 RA Arntmann J., Rutan A.J., Pidsley S.C., Watt F.M., Rees D.A.,
 RA Buxton R.S., Magee A.I.;
 RT Desmosomal glycoprotein Dsg1, a component of intercellular desmosome

RT junctions, is related to the cadherin family of cell adhesion
RT molecules.";
RT Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL AND ESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
(POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56654; CAA39976.1; -
DR PIR: S16906; IJHUG1.
DR HSSP: P09803; IEDR.
DR MIM: 125670; -
DR PROSITE: PS00232; CADHERIN. 2.
DR PFAM: PF00028; cadherin; 4.
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
FT SIGNAL 1 23
FT PROPEP 24 49
FT CHAIN 50 1049
FT DOMAIN 50 545
FT TRANSMEM 546 570
FT DOMAIN 571 1049
FT REPEAT 50 158
FT REPEAT 159 270
FT REPEAT 271 385
FT REPEAT 386 497
FT REPEAT 813 839
FT REPEAT 840 869
FT REPEAT 870 899
FT REPEAT 900 927
FT REPEAT 928 956
FT DOMAIN 969 1019
FT CARBOHYD 36 36
FT CARBOHYD 110 110
FT CARBOHYD 180 180
SQ SEQUENCE 1049 AA: 113715 MW: 113715 MW: EEE125655B9D6619 CRG64;

Query Match 80.0%; Score 84; DB 1; Length 1049;
Best Local Similarity 60.0%; Pred. No. 4.29e-06;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 YGIFVINKTGEINI 111
OY 1 FGIFVVKNTGDINI 15

RESULT 4
ID DSG2_HUMAN STANDARD: PRT: 1117 AA.
AC 014126;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 2 PRECURSOR (HDGC).
GN DSG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON CARCINOMA:
RX MEDLINE: 94192736.

RA Schaefer S., Koch P.J., Franke W.W.;
RT "Identification of the ubiquitous human desmoglein, Dsg2, and the
RT expression catalogue of the desmoglein subfamily of desmosomal
RT cadherins.";
RL Exp. Cell Res. 211:391-399(1994).
RN [2]
RP SEQUENCE OF 777-1117 FROM N.A.
RX MEDLINE: 92037656.
RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R., Franke W.W.;
RT "Complete amino acid sequence of the epidermal desmoglein precursor
RT polypeptide and identification of a second type of desmoglein gene.";
RL Eur. J. Cell Biol. 55:200-208(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
(POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z26317; CAA81226.1; -
DR HSSP: P15116; INCI.
DR MIM: 125671; -
DR PRINTS: PR00205; CADHERIN.
DR PROSITE: PS00232; CADHERIN. 3.
DR PFAM: PF00028; cadherin; 4.
KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KW Cytoskeleton; Calcium-binding.
FT SIGNAL 1 23
FT PROPEP 24 48
FT CHAIN 49 1117
FT DOMAIN 49 608
FT TRANSMEM 609 633
FT DOMAIN 634 1117
FT REPEAT 49 159
FT REPEAT 160 272
FT REPEAT 273 387
FT REPEAT 388 502
FT REPEAT 880 911
FT REPEAT 912 941
FT REPEAT 942 967
FT REPEAT 968 991
FT REPEAT 992 1020
FT REPEAT 1021 1050
FT CARBOHYD 111 111
FT CARBOHYD 181 181
FT CARBOHYD 308 308
FT CARBOHYD 461 461
FT CARBOHYD 513 513
SQ SEQUENCE 1117 AA: 122385 MW: 223857FED70B289 CRG64;

Query Match 78.1%; Score 82; DB 1; Length 1117;
Best Local Similarity 60.0%; Pred. No. 1.28e-05;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 98 FGIFVFKNDGELNV 112
OY 1 FGIFVVKNTGDINI 15

RESULT 5
ID DSC1_BOVIN STANDARD: PRT: 893 AA.
AC 00107; 028095;
DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, last sequence update)
 DT 15-FEB-2000 (Rel. 39, last annotation update)
 DE DESMOCOLLIN 1A/1B PRECURSOR (DESMOSOMAL GLYCOPROTEIN 2/3) (DG2/DG3).
 GN DSCI.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN (1)
 RP SEQUENCE FROM N.A. (1B), AND PARTIAL SEQUENCE.
 RC TISSUE-EPIDERMIS;
 RX MEDLINE: 91185414.
 RA Collins J.E., Legan P.K., Kenny T.P., Macgavyle J., Holton J.L.,
 RT Garrod D.R.;
 RT "Cloning and sequence analysis of desmosomal glycoproteins 2 and 3
 RT (desmocollins): cadherin-like desmosomal adhesion molecules with
 RT heterogeneous cytoplasmic domains.";
 RL J. Cell Biol. 113:381-391(1991).
 RN (2)
 RP SEQUENCE FROM N.A. (1B), AND PARTIAL SEQUENCE.
 RX MEDLINE: 91239591.
 RA Mechanic S., Raynor K., Hill J.E., Cowin P.;
 RT "Desmocollins form a distinct subset of the cadherin family of cell
 RT adhesion molecules.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4476-4480(1991).
 RN (3)
 RP SEQUENCE OF 133-893 FROM N.A. (1A), AND PARTIAL SEQUENCE.
 RC TISSUE-MUZZLE EPITHELIUM;
 RX MEDLINE: 92008912.
 RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R., Schmelz M.,
 RT Franke W.W.;
 RT Amino acid sequence of bovine muzzle epithelial desmocollin derived
 RT from cloned cDNA: a novel subtype of desmosomal cadherins.";
 RL Differentiation 47:29-36(1991).
 RN (4)
 RP SEQUENCE OF 133-155.
 RX MEDLINE: 91115997.
 RA Holton J.L., Kenny T.P., Legan P.K., Collins J.E., Keen J.N.,
 RT Sharma R., Garrod D.R.;
 RT "Desmosomal glycoproteins 2 and 3 (desmocollins) show N-terminal
 RT similarity to calcium-dependent cell-cell adhesion molecules.";
 RL J. Cell Sci. 97:239-246(1990).
 RN (5)
 RP PHOSPHORYLATION.
 RX MEDLINE: 91009551.
 RA Parfitt E.P., Marston J.E., Matley D.L., Measures H.R., Vennart R.,
 RT Garrod D.R.;
 RT "Size heterogeneity, phosphorylation and transmembrane organisation
 RT of desmosomal glycoproteins 2 and 3 (desmocollins) in MDCK cells.";
 RL J. Cell Sci. 96:239-248(1990).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED
 CC TO THE KERATINIZATION OF EPITHELIAL TISSUES.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS: 1A/DG2 (SHOWN HERE) AND 1B/DG3;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EPIDERMIS AND WEAKLY IN TONGUE PAPILLAE.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- PTM: FORM 1A IS PHOSPHORYLATED ON A SERINE BUT FORM 1B IS NOT.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).

DR EMBL: M67489; AAA30492.1; -
 DR EMBL: X56966; CAA40286.1; -
 DR EMBL: X56967; CAA40287.1; -
 DR EMBL: X56968; CAA40289.1; -
 DR EMBL: X56968; CAA40288.1; -
 DR EMBL: X58029; CAA41088.1; -
 DR PIR: B38456; IYBDE.
 DR PIR: A43838; IYBDE.
 DR PIR: A39377; A39377.
 DR HSP: P09803; 1EDH.
 DR PRINTS: P00205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN; 2.
 DR PFAM: PF00028; cadherin; 5.
 KM Cell adhesion: Glycoprotein: Phosphorylation: Transmembrane: Signal;
 KW Repeat: Cytoskeleton: Calcium-binding; Alternative splicing.
 FT SIGNAL 1 29
 FT PROPEP 30 132
 FT CHAIN 133 893
 FT DOMAIN 133 692
 FT TRANSMEM 693 715
 FT DOMAIN 716 893
 FT REPEAT 133 240
 FT REPEAT 241 352
 FT REPEAT 353 470
 FT REPEAT 471 574
 FT REPEAT 575 682
 FT CARBOHYD 163 163
 FT CARBOHYD 398 398
 FT CARBOHYD 545 545
 FT VARIANT 519 519
 FT VARIANT 788 788
 FT VARSPLIC 829 839
 FT VARSPLIC 840 893
 FT CONFLICT 485 485
 FT SEQUENCE 893 AA; 99647 MW; A45A4DBB30951FC9 CRC64;
 SO
 Query Match 62.9%; Score 66; DB 1; Length 893;
 Best Local Similarity 53.8%; Pred. No. 5,19e-02;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 180 FNLFIEKDTGDI 192
 Qy 1 FGIFVVDKNTGDI 13
 RESULT 6
 ID DSCI.HUMAN STANDARD: PRT: 894 AA.
 AC 008554;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE DESMOCOLLIN 1A/1B PRECURSOR (DESMOSOMAL GLYCOPROTEIN 2/3) (DG2/DG3).
 GN DSCI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE-FORSKIN;
 RX MEDLINE: 93283249.
 RA Theis D.G., Koch P.J., Franke W.W.;
 RT "Differential synthesis of type 1 and type 2 desmocollin mRNAs in
 RT human stratified epithelia.";
 RL Int. J. Dev. Biol. 37:101-110(1993).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC TISSUE-FORSKIN;
 RX Zimbelmann R.;
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC TISSUE-SKIN;

RX MEDLINE: 94116981.
 RA King I.A., Arremann J., Spurr N.K., Buxton R.S.:
 RT "Cloning of the cDNA (DSCL) coding for human type 1 desmocollin and
 its assignment to chromosome 18.";
 RL Genomics 18:185-194(1993).
 RN (4)
 RP SEQUENCE OF 135-151 AND 283-292.
 RX MEDLINE: 91323543.
 RA King I.A., Magee A.I., Rees D.A., Buxton R.S.:
 RT "Keratinization is associated with the expression of a new protein
 related to the desmosomal cadherin Dscl/Iti.";
 RL FEBS Lett. 286:9-12(1991).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
 IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED
 TO THE KERATINIZATION OF EPITHELIAL TISSUES.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS: 1A/DG2 (SHOWN HERE) AND 1B/DG3.
 CC -1- ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN EPIDERMIS, LESS IN LYMPH
 NODE AND TONGUE.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 234522; CAA84279.1; -
 DR EMBL: 234522; CAA84278.1; -
 DR EMBL: X72925; CAA51428.1; -
 DR EMBL: X72925; CAA51429.1; -
 DR HSSP: P09803; 1EDH.
 DR MIM: 125643; -
 DR PRINTS: PR00205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN; 3.
 DR PFAM: PF00028; cadherin; 5.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
 KM Cytoskeleton; Calcium-binding; Alternative splicing.
 FT SIGNAL 1 29
 FT PROPEP 30 134
 FT CHAIN 135 884
 FT DOMAIN 135 691
 FT TRANSMEM 692 714
 FT DOMAIN 715 894
 FT REPEAT 135 242
 FT REPEAT 243 354
 FT REPEAT 355 471
 FT REPEAT 472 575
 FT REPEAT 576 682
 FT CARBOHYD 165 165
 FT CARBOHYD 546 546
 FT VARSPLIC 830 840
 FT FT
 FT VARSPLIC 841 894
 FT CONFLICT 132 132
 FT SEQUENCE 894 AA; 100044 MW; 44BA33038699E3E1 CRC64;
 SQ
 Query Match 61.9%; Score 65; DB 1; Length 894;
 Best Local Similarity 53.8%; Pred. No. 8,44e-02;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

RESULT 7
 ID DSCL MOUSE STANDARD: PRT: 886 AA.
 AC P55849;
 DT 01-NOV-1997 (rel. 35, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 15-JUL-1998 (rel. 36, Last annotation update)
 DE DESMOCOLLIN 1A/1B PRECURSOR.
 GN DSCL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6; TISSUE-SKIN.
 RX MEDLINE: 96420658.
 RA King I.A., O'Brien T.J., Buxton R.S.:
 RT "Expression of the 'skin-type' desmosomal cadherin Dscl is closely
 linked to the keratinization of epithelial tissues during mouse
 development.";
 RL J. Invest. Dermatol. 107:531-538(1996).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
 IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED
 TO THE KERATINIZATION OF EPITHELIAL TISSUES.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: FORMS 1A AND 1B ARE PRODUCED BY ALTERNATIVE
 SPLICING OF THE SAME GENE.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X97986; CAA66628.1; -
 DR EMBL: X97986; CAA66629.1; -
 DR HSSP: P09803; 1EDH.
 DR MGD: MGI:109173; DSCI.
 DR PRINTS: PR00205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN; 2.
 DR PFAM: PF00028; cadherin; 5.
 KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
 KM Calcium-binding; Repeat; Alternative splicing.
 FT SIGNAL 1 29
 FT PROPEP 30 134
 FT CHAIN 135 886
 FT DOMAIN 135 691
 FT TRANSMEM 692 714
 FT DOMAIN 715 886
 FT REPEAT 135 242
 FT REPEAT 243 354
 FT REPEAT 355 471
 FT REPEAT 472 575
 FT REPEAT 576 682
 FT CARBOHYD 130 130
 FT CARBOHYD 165 165
 FT CARBOHYD 546 546
 FT CARBOHYD 613 613
 FT VARSPLIC 822 832
 FT FT
 FT VARSPLIC 840 886
 FT SEQUENCE 886 AA; 98953 MW; F34F8D8578CE92F7 CRC64;
 SQ
 Query Match 58.1%; Score 61; DB 1; Length 886;
 Best Local Similarity 46.2%; Pred. No. 5.62e-01;

Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

DB 182 YNLFYKEDTGD 194
 OY 1 FGIVVDKNTGDI 13

RESULT 8
 ID TP2B_MOUSE STANDARD: PRT: 1612 AA.
 AC 064511:
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
 GN TOP2B.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE-BRAIN;
 RA Miyake M., Adachi N., Kikuchi A.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
 CC MAKES DOUBLE-STRAND BREAKS.
 CC CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
 CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
 CC RELAX ONLY NEGATIVE SUPERCOILS.
 CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: D38046; BAA07236.1; -.
 DR HSSP: P06786; 1BGW.
 DR MGD: MGI:98791; TOP2B.
 DR PRINTS: PRO0418; TP12FAMILY.
 DR PRINTS: PRO0615; CCAATSUBUNTA.
 DR PRINTS: PRO1158; TOPISMRASEII.
 DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
 DR PFMW: PFM0204; DNA_topoisoi1; 1.
 DR NP_BIND 170 175 ATP (POTENTIAL).
 FT ACT_SITE 814 814 DNA CLEAVAGE (BY SIMILARITY).
 FT ACT_SITE 1612 AA; 181863 MW; 331B930065130864 CRC64;
 SO SEQUENCE

Query Match 58.18; Score 61; DB 1; Length 1612;
 Best Local Similarity 72.78; Pred. NO. 5.62e-01;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 922 IFVVDRTVEI 932
 OY 3 IFVVDKNTGDI 13

RESULT 9
 ID TP2B_CRILLO STANDARD: PRT: 1612 AA.
 AC 064399:
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
 GN TOP2B.

OS Cricetus longicaudatus (long-tailed hamster) (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LUNG;
 RX MEDLINE: 96085121.
 RA Derouddre S., Frey S., Delaportie C., Jacquemin-Sablon A.;
 RT "Cloning and characterization of full-length cDNAs coding for the DNA
 RT topoisomerase II beta from Chinese hamster lung cells sensitive and
 RT resistant 9-OH-ellipticine."
 RL Biochim. Biophys. Acta 1264:178-182(1995).
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II;
 CC MAKES DOUBLE-STRAND BREAKS.
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
 CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
 CC RELAX ONLY NEGATIVE SUPERCOILS.
 CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X86455; CAA60173.1; -.
 DR HSSP: P06786; 1BGW.
 DR PRINTS: PRO0418; TP12FAMILY.
 DR PRINTS: PRO0615; CCAATSUBUNTA.
 DR PRINTS: PRO1158; TOPISMRASEII.
 DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
 DR PFMW: PFM0204; DNA_topoisoi1; 1.
 DR NP_BIND 170 175 ATP (POTENTIAL).
 FT ACT_SITE 814 814 DNA CLEAVAGE (BY SIMILARITY).
 FT ACT_SITE 1612 AA; 182074 MW; C01D6FC40620FC68 CRC64;
 SO SEQUENCE

Query Match 58.18; Score 61; DB 1; Length 1612;
 Best Local Similarity 72.78; Pred. NO. 5.62e-01;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 922 IFVVDRTVEI 932
 OY 3 IFVVDKNTGDI 13

RESULT 10
 ID TP2B_HUMAN STANDARD: PRT: 1626 AA.
 AC 002880:
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
 GN TOP2B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 93087155.
 RA Jenkins J.R., Ayton P., Jones T., Davies S.L., Simmons D.L.,
 RA Harris A.L., Sheer D., Hickson I.D.;
 RT "Isolation of cDNA clones encoding the beta isozyme of human DNA
 RT topoisomerase II and localisation of the gene to chromosome 3p24."
 RL Nucleic Acids Res. 20:5587-5592(1992).
 RN [2]

RP. SEQUENCE FROM N.A.
 RX MEDLINE: 93192319.
 RA Austin C.A., Sing J.H., Patel S., Fisher L.M.:
 RT "Novel Hela topoisomerase II is the II beta isoform: complete coding
 RL sequence and homology with other type II topoisomerases.";
 RN Biochim. Biophys. Acta 1172:283-291(1993).
 [3]
 RP SEQUENCE OF 1038-1271 FROM N.A.
 RX MEDLINE: 90306333.
 RA Austin C.A., Fisher L.M.:
 RT "Isolation and characterization of a human cDNA clone encoding a
 RL novel DNA topoisomerase II homologue from Hela cells.";
 RN FEBS Lett. 266:115-117(1990).
 [4]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE: 93376494.
 RA Davies S.L., Jenkins J.R., Hickson I.D.:
 RT "Human cells express two differentially spliced forms of
 RL topoisomerase II beta mRNA.";
 RN Nucleic Acids Res. 21:3719-3723(1993).
 CC -I- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
 CC MAKES DOUBLE-STRAND BREAKS.
 CC -I- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 CC -I- SUBUNIT: HOMODIMER.
 CC -I- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
 CC -I- ALTERNATIVE PRODUCTS: TWO FORMS, CALLED BETA-1 AND BETA-2 ARE
 CC PRODUCED BY ALTERNATIVE SPLICING OF THE TOP2B GENE. THE BETA-2
 CC FORM IS SHOWN HERE.
 CC -I- PTM: PHOSPHORYLATED.
 CC -I- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
 CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
 CC RELAX ONLY NEGATIVE SUPERCOILS.
 CC -I- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 DR EMBL: X68060; CAA48197.1; -;
 DR EMBL: J71911; -; NOT ANNOTATED_CDS.
 DR EMBL: 215111; CAA78815.1; -;
 DR EMBL: 215115; CAA78821.1; -;
 DR EMBL: X53662; CAA37706.1; -;
 DR PIR: S26730; S26730.
 DR PIR: S10710; S10710.
 DR PIR: S41641; S41641.
 DR HSSP: P06786; IBCW.
 DR MIM: 126431.
 DR PRINTS: PR00418; TP12FAMILY.
 DR PRINTS: PR00615; CCAATSUBUNTA.
 DR PRINTS: PR01158; TOPISMRASEII.
 DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
 DR PFAM: PF00204; DNA_topoisomII; 1.
 KW Isomerase: Topoisomerase: DNA-binding; ATP-binding; Phosphorylation;
 KW Alternative splicing; Nuclear protein.
 KM NP_BIND 182 187 ATP (POTENTIAL).
 FT ACT_SITE 826 826 DNA CLEAVAGE (BY SIMILARITY).
 FT VARSPIC 24 28 MISSING (IN ISOFORM BETA-1).
 FT CONFLICT 1611 1611 T -> A (IN REF. 2).
 SQ SEQUENCE 1626 AA; 183296 MW; E7BE262CC68B04D CRC64;
 Query Match 58.1%; Score 61; DB 1; Length 1626;
 Best Local Similarity 72.7%; Pred. No. 5,62e-01;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 DB 934 IFVVDKNTVEI 944
 |||||:|:|:|

OY 3 IFVVDKNTGDI 13
 RESULT 11
 ID TP2B_CHICK STANDARD; PRT: 1627 AA.
 AC 042131;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
 GN TOP2B.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
 OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nimi A., Harata M., Mizuno S.:
 RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
 CC -I- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
 CC MAKES DOUBLE-STRAND BREAKS (BY SIMILARITY).
 CC -I- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 CC -I- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -I- SUBCELLULAR LOCATION: NUCLEAR; GENERALLY LOCATED IN THE
 CC NUCLEOLUS.
 CC -I- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
 CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
 CC RELAX ONLY NEGATIVE SUPERCOILS.
 CC -I- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 DR EMBL: AB007446; BAA2540.1; -;
 DR HSSP: P06786; IBCW.
 DR PRINTS: PR00418; TP12FAMILY.
 DR PRINTS: PR00615; CCAATSUBUNTA.
 DR PRINTS: PR01158; TOPISMRASEII.
 DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
 DR PFAM: PF00204; DNA_topoisomII; 1.
 KW Isomerase: Topoisomerase: DNA-binding; ATP-binding; Nuclear protein.
 KM NP_BIND 187 192 ATP (POTENTIAL).
 FT ACT_SITE 831 831 DNA CLEAVAGE (BY SIMILARITY).
 FT DOMAIN 5 23 POLY-GLY.
 FT DOMAIN 1265 1268 POLY-LYS.
 FT DOMAIN 1388 1391 POLY-ASP.
 FT DOMAIN 1393 1396 POLY-ASN.
 SQ SEQUENCE 1627 AA; 183245 MW; 8B651D10A2C4D34B CRC64;
 Query Match 58.1%; Score 61; DB 1; Length 1627;
 Best Local Similarity 72.7%; Pred. No. 5,62e-01;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 DB 939 IFVVDKNTVEI 949
 |||||:|:|:|
 OY 3 IFVVDKNTGDI 13
 RESULT 12
 ID YRHB_ECOLI STANDARD; PRT: 94 AA.
 AC P46857;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL 10.6 KD PROTEIN IN GNT-R-GGT INTERGENIC REGION (094).
 GN YRHB.
 OS Escherichia coli.

OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Colado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.D.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U18997; AAA58244.1; -
 CC EMBL: AE000421; AAC76471.1; -
 CC ECGENE: EGI2948; yrb.
 CC Hypothetical protein.
 KM SEQUENCE 94 AA; 10613 MW; 8A86265E70E5B06 CRC64;
 SO
 Query Match 57.1%; Score 60; DB 1; Length 94;
 Best Local Similarity 45.3%; Pred. No. 8,91e-01;
 Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
 DB 59 FIIDKSGEIH 69
 ID 4 FVVDKNTGDI 14
 QY
 RESULT 13
 ID CADE_HUMAN STANDARD; PRT; 790 AA.
 AC Q13634;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CADHERIN-14 PRECURSOR.
 GN CDH14.
 OS Homo sapiens (Human).
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
 OC Eutheria: Primates: Catarrhini: Homnidae: Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 97184182.
 RX Shibata T., Shimoyama Y., Gotoh M., Hirohashi S.;
 RA Identification of human cadherin-14, a novel neurally specific type
 RT I cadherin, by protein interaction cloning.";
 RL J. Biol. Chem. 272:5226-5240(1997).
 CC -1- FUNCTION: CADHERINS ARE CALCINUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES.
 CC SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U59325; AAB02933.1; -
 CC HSSP: P15116; INCI.
 DR PRINTS: PR00205; CADHERIN.

DR PROSITE: PS00232; CADHERIN; 3.
 DR PFAM: PF00028; cadherin; 5.
 DR PFAM: PF01049; Cadherin_C-term; 1.
 DR Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
 KM Calcium-binding; Repeat; Signal.
 FT SIGNAL 1 24
 FT PROPEP 25 53
 FT CHAIN 54 790
 FT DOMAIN 54 608
 FT TRANSMEM 609 636
 FT DOMAIN 637 790
 FT REPEAT 54 159
 FT REPEAT 160 268
 FT REPEAT 269 383
 FT REPEAT 384 486
 FT REPEAT 487 608
 FT CARBOHYD 36 36
 FT CARBOHYD 255 255
 FT CARBOHYD 455 455
 FT CARBOHYD 536 536
 SO SEQUENCE 790 AA; 86072 MW; 5C7BDEB229B6EDCA CRC64;
 Query Match 57.1%; Score 60; DB 1; Length 790;
 Best Local Similarity 58.3%; Pred. No. 8,91e-01;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 DB 100 IFIIDTGTGDI 111
 ID 3 IFVVDKNTGDI 14
 QY
 RESULT 14
 ID POLG_PYEVI STANDARD; PRT; 3027 AA.
 AC Q05057;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GENOME POLYPROTEIN [CONTAINS: 22.5 KD PROTEIN; 26 KD PROTEIN; 31 KD
 DE PROTEIN; PROBABLE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48)].
 OS Parsnip yellow fleck virus (isolate P-121) (PYFV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Sequiviridae;
 OC Sequivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93107855.
 RA Turnbull-Ross A.D., Reavy B., Mayo M.A., Murrant A.F.;
 RT The nucleotide sequence of parsnip yellow fleck virus: a plant
 RT picorna-like virus.";
 RL J. Gen. Virol. 73:3203-3211(1992).
 CC -1- SIMILARITY: SOME TO THE CMV AND TBV POLYPROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D14066; BAA03151.1; -
 DR PIR: J01917; J01917.
 DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
 DR PFAM: PF00910; RNA_helicase; 1.
 KM Polyprotein; ATP-binding; Coat protein; Transferase;
 KW RNA-directed RNA polymerase.
 FT NP_BIND 1467 1474
 FT VARIANT 962 962
 FT VARIANT 1373 1373
 FT VARIANT 1373 1373
 SO SEQUENCE 3027 AA; 336242 MW; 0C41EB985F405BE2 CRC64;
 Query Match 55.2%; Score 58; DB 1; Length 3027;
 Best Local Similarity 40.0%; Pred. No. 2,21e-00;
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch.p protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 07:00:24 2000; Maspar time 232.53 Seconds

Tabular output not generated. 4.473 Million cell updates/sec

Title: >US-08-991-628-2

Description: (1-15) from US08991628.pap

Perfect Score: 105

Sequence: 1 FGIFVVDKNTGDIINI 15

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

sptrembl12
 1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
 5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
 9:sp_phage 10:sp_plant 11:sp_ricent 12:sp_unclassified
 13:sp_vertebrate 14:sp_virus

Statistics: Mean 26.732; Variance 34.969; scale 0.764

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|------------|-------|-------------|--------|--------|------------------------|-------------------------|
| 1 | 90 | 85.7 | 993 | 11 | 035902 | DESMOGLEIN 3 (FRAGMENT) |
| 2 | 66 | 62.9 | 785 | 13 | 020763 | CHICKEN CADHERIN-7. |
| 3 | 63 | 60.0 | 1329 | 5 | 076356 | C45G7.6 PROTEIN. |
| 4 | 61 | 58.1 | 3503 | 5 | 024292 | ADHERIN. |
| 5 | 60 | 57.1 | 214 | 2 | 055701 | HYPOTHETICAL 24.5 KD P |
| 6 | 57.1 | 598 | 1 | 029907 | PROBABLE TUNGSTEN-CONT | |
| 7 | 56.2 | 154 | 1 | 058735 | HYPOTHETICAL PROTEIN M | |
| 8 | 55.2 | 245 | 1 | 058723 | HYPOTHETICAL PROTEIN M | |
| 9 | 55 | 56.2 | 624 | 10 | 048662 | PV72. |
| 10 | 59 | 56.2 | 1033 | 2 | 033373 | PILC PROTEIN. |
| 11 | 58 | 55.2 | 270 | 1 | 028514 | CONSERVED HYPOTHETICAL |
| 12 | 58 | 55.2 | 555 | 10 | 0920W3 | LACCASE (EC 1.10.3.2). |
| 13 | 58 | 55.2 | 593 | 10 | 0920W3 | HYPOTHETICAL 68.6 KD P |
| 14 | 58 | 55.2 | 816 | 4 | 075284 | KIAA0345-LIKE 9. |
| 15 | 58 | 55.2 | 936 | 4 | 0955H7 | PROTODADHERIN ALPHA 5. |
| 16 | 58 | 55.2 | 950 | 13 | 095508 | PARAXIAL PROTODADHERIN |
| 17 | 58 | 55.2 | 1035 | 13 | 057537 | NE-PROTODADHERIN |
| 18 | 58 | 55.2 | 1069 | 4 | 060245 | PCDH7 (BH-PCDH)A. |
| 19 | 58 | 55.2 | 1072 | 4 | 060246 | PCDH7 (BH-PCDH)B. |
| 20 | 58 | 55.2 | 1200 | 4 | 060247 | PCDH7 (BH-PCDH)C. |

| 21 | 57 | 54.3 | 209 | 2 | 092JRI | RIBONUCLEASE HII. | 7.66e+00 |
|----|----|--|------|----|--------|------------------------|----------|
| 22 | 57 | 54.3 <td>504</td> <td>1</td> <td>057984</td> <td>504AA LONG HYPOTHETICA</td> <td>7.66e+00</td> | 504 | 1 | 057984 | 504AA LONG HYPOTHETICA | 7.66e+00 |
| 23 | 57 | 54.3 <td>693</td> <td>4</td> <td>015066</td> <td>OB-CADHERIN-2.</td> <td>7.66e+00</td> | 693 | 4 | 015066 | OB-CADHERIN-2. | 7.66e+00 |
| 24 | 57 | 54.3 <td>792</td> <td>13</td> <td>093319</td> <td>CADHERIN 11.</td> <td>7.66e+00</td> | 792 | 13 | 093319 | CADHERIN 11. | 7.66e+00 |
| 25 | 57 | 54.3 <td>794</td> <td>13</td> <td>093264</td> <td>CADHERIN PRECURSOR.</td> <td>7.66e+00</td> | 794 | 13 | 093264 | CADHERIN PRECURSOR. | 7.66e+00 |
| 26 | 57 | 54.3 <td>797</td> <td>11</td> <td>063418</td> <td>PUTATIVE LIPOPROTEIN.</td> <td>7.66e+00</td> | 797 | 11 | 063418 | PUTATIVE LIPOPROTEIN. | 7.66e+00 |
| 27 | 57 | 54.3 <td>798</td> <td>2</td> <td>050288</td> <td>PUTATIVE LIPOPROTEIN.</td> <td>7.66e+00</td> | 798 | 2 | 050288 | PUTATIVE LIPOPROTEIN. | 7.66e+00 |
| 28 | 57 | 54.3 <td>803</td> <td>4</td> <td>075283</td> <td>KIAA0345-LIKE 8.</td> <td>7.66e+00</td> | 803 | 4 | 075283 | KIAA0345-LIKE 8. | 7.66e+00 |
| 29 | 57 | 54.3 <td>814</td> <td>4</td> <td>075281</td> <td>KIAA0345-LIKE 6.</td> <td>7.66e+00</td> | 814 | 4 | 075281 | KIAA0345-LIKE 6. | 7.66e+00 |
| 30 | 57 | 54.3 <td>950</td> <td>4</td> <td>0915H6</td> <td>PROTODADHERIN ALPHA 8.</td> <td>7.66e+00</td> | 950 | 4 | 0915H6 | PROTODADHERIN ALPHA 8. | 7.66e+00 |
| 31 | 56 | 53.3 <td>171</td> <td>5</td> <td>061171</td> <td>CENTRIN 1 (FRAGMENT).</td> <td>1.18e+01</td> | 171 | 5 | 061171 | CENTRIN 1 (FRAGMENT). | 1.18e+01 |
| 32 | 56 | 53.3 <td>205</td> <td>5</td> <td>027342</td> <td>GLUTATHIONE PEROXIDASE</td> <td>1.18e+01</td> | 205 | 5 | 027342 | GLUTATHIONE PEROXIDASE | 1.18e+01 |
| 33 | 56 | 53.3 <td>423</td> <td>10</td> <td>065875</td> <td>CELL DIVISION PROTEIN</td> <td>1.18e+01</td> | 423 | 10 | 065875 | CELL DIVISION PROTEIN | 1.18e+01 |
| 34 | 56 | 53.3 <td>601</td> <td>2</td> <td>086691</td> <td>PUTATIVE TRANSPORT SYS</td> <td>1.18e+01</td> | 601 | 2 | 086691 | PUTATIVE TRANSPORT SYS | 1.18e+01 |
| 35 | 56 | 53.3 <td>649</td> <td>10</td> <td>09XFN8</td> <td>17F8.5</td> <td>1.18e+01</td> | 649 | 10 | 09XFN8 | 17F8.5 | 1.18e+01 |
| 36 | 56 | 53.3 <td>788</td> <td>4</td> <td>0916N8</td> <td>CADHERIN-10.</td> <td>1.18e+01</td> | 788 | 4 | 0916N8 | CADHERIN-10. | 1.18e+01 |
| 37 | 56 | 53.3 <td>831</td> <td>5</td> <td>P91388</td> <td>COSMID K1209.</td> <td>1.18e+01</td> | 831 | 5 | P91388 | COSMID K1209. | 1.18e+01 |
| 38 | 55 | 52.4 | 143 | 2 | 066117 | HYPOTHETICAL 16.3 KD P | 1.18e+01 |
| 39 | 55 | 52.4 | 261 | 13 | 042389 | SUCRASE-ISOMALTASE (PR | 1.18e+01 |
| 40 | 55 | 52.4 | 340 | 10 | 022706 | F8A5.20 PROTEIN. | 1.18e+01 |
| 41 | 55 | 52.4 | 569 | 14 | 069013 | POLYMERASE (FRAGMENT). | 1.18e+01 |
| 42 | 55 | 52.4 | 706 | 13 | 093869 | GLYCOGEN SYNTHASE | 1.18e+01 |
| 43 | 55 | 52.4 | 790 | 13 | 090762 | CHICKEN CADHERIN-6B. | 1.18e+01 |
| 44 | 55 | 52.4 | 1196 | 2 | 045916 | 138RDA PROTEIN ASSOCIA | 1.18e+01 |
| 45 | 55 | 52.4 | 1196 | 9 | 092X77 | NTNH. | 1.18e+01 |

ALIGNMENTS

| RESULT | 1 | PREDIMINARY | PRT | 993 AA. |
|--------|--|-------------|-----|---------|
| ID | 035902 | | | |
| AC | 035902 | | | |
| DT | 01-JAN-1998 (TREMBLrel. 05, Created) | | | |
| DT | 01-JAN-1998 (TREMBLrel. 05, Last sequence update) | | | |
| DT | 01-NOV-1999 (TREMBLrel. 12, Last annotation update) | | | |
| DE | DESMOGLEIN 3 (FRAGMENT). | | | |
| GN | DSG3. | | | |
| OS | Mus musculus (Mouse). | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; | | | |
| NC | Utheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RA | STRAIN-BALB/C; | | | |
| IS | ISHIKAWA H., LI K., UIRTO J.; | | | |
| RL | Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases. | | | |
| CC | -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY). | | | |
| DR | EMBL; U86016; AAB5091.1; -. | | | |
| DR | HSSP; P15116; INCU. | | | |
| DR | PROSITE; PS00232; CADHERIN; 2. | | | |
| DR | PFAM; PF00028; cadherin; 4. | | | |
| KW | Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat. | | | |
| FT | NON-TER 993 | | | |
| SO | SEQUENCE 993 AA; 10788 MW; 881794BD CRC32; | | | |

Query Match 85.7%; Score 90; DB 11; Length 993;
 Best Local Similarity 86.7%; Pred. No. 4.79e+07;
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

| Db | 97 | FGIFVVDKNTGDIINI 11 |
|--------|--|---------------------|
| Oy | 1 | FGIFVVDKNTGDIINI 15 |
| RESULT | 2 | |
| ID | 090763 | PRELIMINARY; |
| AC | 090763 | PRT; 785 AA. |
| DT | 01-NOV-1996 (TREMBLrel. 01, Created) | |
| DT | 01-NOV-1996 (TREMBLrel. 01, Last sequence update) | |
| DT | 01-NOV-1999 (TREMBLrel. 12, Last annotation update) | |
| DE | CHICKEN CADHERIN-7. | |
| OS | Gallus gallus (Chicken). | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves; | |
| NC | Neognathae; Galliformes; Phasianidae; Phasianidae; Gallus. | |

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WHITE LECHORN: TISSUE-BRAIN:
 RX MEDLINE: 95309115.
 RA NAKAGAWA S., TAKEICHI M.:
 RT "Neural crest cell-cell adhesion controlled by sequential and
 subpopulation-specific expression of novel cadherins."
 RL Development 121:1321-1332(1995).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: D42150; BAA07721.1; -.
 DR HSSP: P15116; INCB.
 DR PROSITE: PS00232; CADHERIN: 3.
 DR PFAM: PF00028; cadherin: 5.
 DR PFAM: PF01049; Cadherin_C-term: 1.
 DR PRINTS: PR00205; CADHERIN.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SQ SEQUENCE 785 AA: 87177 MW: F27A6881 CRC32:

Query Match 62.9%; Score 66; DB 13; Length 785;
 Best Local Similarity 61.5%; Pred. No. 1.24e-01;
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 93 SIFIDENTGDIH 105
 Qy 2 GIFVVDKNTGDI 14

RESULT 3 PRELIMINARY: PRT: 1329 AA.
 AC 076356;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE C4567.6 PROTEIN.
 GN C4567.6
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditioidea; Rhabditidae; Pelodermidae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718
 RA WILSON R., AINSFORTH R., ANDERSON K., BAYNES C., BERKS M.,
 BONFIELD J., BORTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 GARDNER A., GREEN P., HAWKINS T., HILLIER L., JER M., JOHNSTON L.,
 JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
 SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RT Nature 368:32-38(1994).
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA DANTE M., WAMBLEY P.;
 RT "The sequence of C. elegans cosmid C4567."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF067611; AAC19184.1; -.
 DR PROSITE: PS00232; CADHERIN: 1.
 DR PFAM: PF00028; cadherin: 3.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SQ SEQUENCE 1329 AA: 146518 MW: B161D39E CRC32:

Query Match 60.0%; Score 63; DB 5; Length 1329;
 Best Local Similarity 72.7%; Pred. No. 5.11e-01;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1288 IFVADKDSGDI 1298
 Qy 3 IFVVDKNTGDI 13

RESULT 4 PRELIMINARY: PRT: 3503 AA.
 AC 024292;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE ADHERIN.
 GN DACHSOUS.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95324813.
 RA CLARK H.F., BRENTNUP D., SCHNEITZ K., BIBER A., GOODMAN C., NOLL M.;
 RT "Dachsous encodes a member of the cadherin superfamily that controls
 imaginal disc morphogenesis in Drosophila."
 RL Genes Dev. 9:1530-1542(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA NOLL M.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: L08811; AAB79329.2; -.
 DR HSSP: P15116; INCB.
 DR FLYBASE: FBgn0000497; ds.
 DR PROSITE: PS00232; CADHERIN: 20.
 DR PFAM: PF00028; cadherin: 25.
 DR PRINTS: PR00205; CADHERIN.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SQ SEQUENCE 3503 AA: 379719 MW: 16030CA5 CRC32:

Query Match 58.1%; Score 61; DB 5; Length 3503;
 Best Local Similarity 58.1%; Pred. No. 1.28e-00;
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 713 GIFRIDRSTGEI 724
 Qy 2 GIFVVDKNTGDI 13

RESULT 5 PRELIMINARY: PRT: 214 AA.
 AC 055701;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
 DE HYPOTHETICAL 24.5 KD PROTEIN.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RA TABATA S.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RX MEDLINE: 96127529.
 RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
 RA SUGIURA M., TABATA S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb

RX MEDLINE: 96337999.
RA BUT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAUGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBEK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
RA SCOTT J.L., GEOCHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA KLENN H.-P., FRASER C.M., SMITH H.O., MOESE C.R., VENTER J.C.,
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
DR EMBL: U67573; AAB99343.1; .
DR TIGR: MJ1327. .
KW Hypothetical protein.
SQ SEQUENCE - 245 AA; 28779 MW; 3D81831F CRC32;

Query Match 56.2%; Score 59; DB 1; Length 245;
Best Local Similarity 87.5%; Pred. No. 3.17e+00;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 151 YGIFVVDK 158
OY 1 FGIFVVDK 8

RESULT 9 PRELIMINARY: PRT: 624 AA.
ID 048662;
AC 048662;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE PVT2.
OS *Cucurbita maxima* (Pumpkin) (Winter squash).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eustosids I; Cucurbitales; Cucurbitaceae;
OC Cucurbita.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-DEVELOPING COTYLEDON;
RX MEDLINE: 98182943.
RA SHIMADA T., KUROYANAGI M., NISHIMURA M., HARA-NISHIMURA I.,
RT "A pumpkin 72-kDa membrane protein of precursor-accumulating vesicles
RL has characteristics of a vacuolar sorting receptor.";
RL Plant Cell Physiol. 38:1414-1420(1997).
DR EMBL: AB006809; BAA25079.1; .
DR HSSP: P35555; JEMO.
DR MENDEL: 24651; Cucum:1391.24651.
DR PROSITE: PS0010; ASX_HYDROXYL; 1.
DR PROSITE: PS01187; EGF_CA; 1.
KW Glycoprotein; EGF-like domain.
SQ SEQUENCE 624 AA; 69028 MW; AAA97D0B CRC32;

Query Match 56.2%; Score 59; DB 10; Length 624;
Best Local Similarity 70.0%; Pred. No. 3.17e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 20 FGFEVVKNS 29
OY 1 FGIFVVDKNT 10

RESULT 10 PRELIMINARY: PRT: 1033 AA.
ID 033373;
AC 033373;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE PILC PROTEIN.
GN PILC.
OS *Neisseria gonorrhoeae*.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GC-653.
RX MEDLINE: 98129088.
RA BACKMAN M., KABELSTROM H., JONSSON A.B.,
RT "The phase-variable plus-associated protein PilC is commonly
RT expressed in clinical isolates of *Neisseria gonorrhoeae*, and shows
RT sequence variability among strains.";
RL Microbiology 144:149-156(1998).
DR EMBL: AJ001121; CA04547.1; .
SQ SEQUENCE 1033 AA; 112936 MW; B0593053 CRC32;

Query Match 56.2%; Score 59; DB 2; Length 1033;
Best Local Similarity 46.7%; Pred. No. 3.17e+00;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 804 YGIFDDKGTGTAVY 818
OY 1 FGIFVVDKNTGDINI 15

RESULT 11 PRELIMINARY: PRT: 270 AA.
ID 028514;
AC 028514;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN AF1760.
OS *Archaeoglobus fulgidus*.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.
RA KLENN H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L.,
RA OVERBEK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., MOESE C.R.,
RA VENTER J.C.,
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon *Archaeoglobus fulgidus*.";
RL Nature 390:364-370(1997).
DR EMBL: AE000981; AAB89491.1; .
DR HSSP: O57997; IMJH.
DR TIGR: AF1760; .
DR PFAM: PF00582; usp; 1.
KW Hypothetical protein.
SQ SEQUENCE 270 AA; 29554 MW; E5FA8A9B CRC32;

Query Match 55.2%; Score 58; DB 1; Length 270;
Best Local Similarity 40.0%; Pred. No. 4.94e+00;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

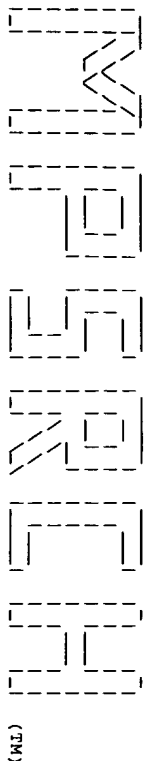
Db 165 YAKFEVKKTGGEIHI 179
OY 1 FGIFVVDKNTGDINI 15

RESULT 12 PRELIMINARY: PRT: 555 AA.
ID 0920W3;
AC 0920W3;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE LACCASE (EC 1.10.3.2).

GN LAC3.
 OS Populus trichocarpa (Western balsam poplar).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 OC eudicotyledons: Rosidae: eudicotyledons: Malvales: Malvaceae: Populus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:CV. TRICHOBL. TISSUE=XYLEM;
 RX MEDLINE: 99115479.
 RA RANOCIA P., MCDUGALL G., HAWKINS S., STERJADES R., BORDERIES G.,
 RA STEWART D., CABANES-MACHEAU M., BOUDER A.M., GOFNER D.,
 RT "Biochemical characterization, molecular cloning and expression of
 RT laccases - a divergent gene family - in poplar."
 RL Eur. J. Biochem. 259:485-495(1999).
 DR EMBL: Y13771; CAA74103.1; -
 DR HSSP: P37064; JASP.
 DR MENDEL: 34836; Popba;1457;34836.
 KM OXIDOREDUCTASE.
 SQ SEQUENCE 555 AA: 60890 MW: 982560EC CRC32:
 Query Match 55.2%; Score 58; DB 10; Length 555;
 Best Local Similarity 42.9%; Pred. No. 4.94e+00;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 Db 463 FNFRVVGKIGNF 476
 QY 1 FGIFVVDKNTGDIN 14
 RESULT 13
 ID 065522 PRELIMINARY: PRT: 593 AA.
 AC 065522;
 DT 01-AUG-1998 (TEMBLrel. 07, Created)
 DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
 DE HYPOHETICAL 68.6 KD PROTEIN.
 GN PAD1.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 OC eudicotyledons: Spermatophyta: Magnoliophyta: eudicotyledons:
 OC core eudicots: Rosidae: eurosids II: Brassicales: Brassicaceae:
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., BENES V., RECHMANN S., BORKOVA D., ANSORGE W., HOHEISEL J.,
 RA MEWES H.W., MAYER K.F.X., SCHUELLER C.;
 RN Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL022537; CAA18582.1; -
 DR MENDEL: 29102; Arch;3426;29102.
 KM Hypochemical protein.
 SQ SEQUENCE 593 AA: 68632 MW: 9EA10B88 CRC32:
 Query Match 55.2%; Score 58; DB 10; Length 593;
 Best Local Similarity 35.7%; Pred. No. 4.94e+00;
 Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
 Db 528 FAVEVLEKIDPLDK 541
 QY 1 FGIFVVDKNTGDIN 14
 RESULT 14
 ID 075284 PRELIMINARY: PRT: 816 AA.
 AC 075284;
 DT 01-NOV-1998 (TEMBLrel. 08, Created)
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
 DE KIA0345-LIKE 9.
 GN PCDH-ALPHA5.

OS Homo sapiens (Human).
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
 OC Eutheria: Primates: Catarrhini: Homiinae: Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA KIMBERLY W., BOND M., CHENG J., CONNOLLY K.S., GUNNING K.M.,
 RA KADNER K., MIGUEL T., MILLER C., PILLUCK S., POLLARD M., ROJESKI H.,
 RA SUBRAMANIAN S., MARTIN C.H.;
 RT "Sequencing of human chromosome 5."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA RICKE D.O.;
 RT "Large Scale Sequence Analysis and Annotation with the Sequence
 RT Comparison Analysis (SCAN) System."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 99308636.
 RA WU Q., MANIATIS T.;
 RT "A striking organization of a large family of human neural cadherin-
 RT like cell adhesion genes."
 RL Cell 97:779-790(1999).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AC005609; AAC34321.1; -
 DR EMBL: AF152483; AAD3744.1; -
 DR PROSITE: PS00232; CADHERIN; 5.
 DR PFAM: PF00028; cadherin; 5.
 DR PRINTS: PR00205; CADHERIN.
 KM Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SQ SEQUENCE 816 AA: 89348 MW: 3F41D941 CRC32:
 Query Match 55.2%; Score 58; DB 4; Length 816;
 Best Local Similarity 41.7%; Pred. No. 4.94e+00;
 Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
 Db 292 FIINSNTGEIKV 303
 QY 4 FVVDKNTGDINI 15
 RESULT 15
 ID 0955H7 PRELIMINARY: PRT: 936 AA.
 AC 0955H7;
 DT 01-NOV-1999 (TEMBLrel. 12, Created)
 DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
 DE PROTOCADHERIN ALPHA 5.
 GN PCDH-ALPHA5.
 OS Homo sapiens (Human).
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
 OC Eutheria: Primates: Catarrhini: Homiinae: Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 99308636.
 RA WU Q., MANIATIS T.;
 RT "A striking organization of a large family of human neural cadherin-
 RT like cell adhesion genes."
 RL Cell 97:779-790(1999).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF152483; AAD3744.1; -
 DR PROSITE: PS00232; CADHERIN; 5.
 KM Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SQ SEQUENCE 936 AA: 102048 MW: 387ECF8D CRC32:
 Query Match 55.2%; Score 58; DB 4; Length 936;
 Best Local Similarity 41.7%; Pred. No. 4.94e+00;
 Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
 Db 292 FIINSNTGEIKV 303
 QY 4 FVVDKNTGDINI 15

Search completed: Sat May 13 07:04:30 2000
Job time : 246 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 06:06:16 2000: Maspar time 52.14 Seconds

Tabular output not generated. 3.727 Million cell updates/sec

Title: >US-08-991-628-2

Description: (1-15) from US08991628.pep

Perfect Score: 105

Sequence: 1 FGIFVVDKNTGDIINI 15

Scoring table: PAM 150

Gap 15

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-issued

1:5A.COMB 2:5B.COMB 3:PCT.COMB 4:backfiles1

Statistics: Mean 17.421; Variance 58.153; scale 0.300

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|------------|-------------------------|-----------|
| 1 | 105 | 100.0 | 15 | 2 | US-08-400- | Sequence 2, Applicatio | 2.01e-04 |
| 2 | 57 | 54.3 | 615 | 2 | US-08-738- | Sequence 12, Applicati | 3.04e+01 |
| 3 | 57 | 54.3 | 693 | 2 | US-08-738- | Sequence 6, Applicatio | 3.04e+01 |
| 4 | 57 | 54.3 | 796 | 2 | US-08-738- | Sequence 2, Applicatio | 3.04e+01 |
| 5 | 57 | 54.3 | 796 | 2 | US-08-188- | Sequence 36, Applicatio | 3.04e+01 |
| 6 | 57 | 54.3 | 796 | 2 | US-08-738- | Sequence 4, Applicatio | 3.04e+01 |
| 7 | 57 | 54.3 | 796 | 1 | US-08-332- | Sequence 58, Applicati | 3.04e+01 |
| 8 | 57 | 54.3 | 796 | 1 | US-08-332- | Sequence 112, Applicat | 3.04e+01 |
| 9 | 57 | 54.3 | 797 | 1 | US-08-268- | Sequence 112, Applicat | 3.04e+01 |
| 10 | 57 | 54.3 | 797 | 2 | US-08-453- | Sequence 112, Applicat | 3.04e+01 |
| 11 | 57 | 54.3 | 797 | 2 | US-08-453- | Sequence 112, Applicat | 3.04e+01 |
| 12 | 57 | 54.3 | 797 | 3 | PCT-US95-0 | Sequence 112, Applicat | 3.04e+01 |
| 13 | 57 | 54.3 | 837 | 2 | US-08-472- | Sequence 6, Applicatio | 3.04e+01 |
| 14 | 57 | 54.3 | 837 | 2 | US-08-474- | Sequence 7, Applicatio | 3.04e+01 |
| 15 | 57 | 54.3 | 837 | 2 | US-08-474- | Sequence 7, Applicatio | 3.04e+01 |
| 16 | 56 | 53.3 | 433 | 2 | US-08-883- | Sequence 2, Applicatio | 3.82e+01 |
| 17 | 56 | 53.3 | 511 | 2 | US-08-537- | Sequence 4, Applicatio | 3.82e+01 |
| 18 | 56 | 53.3 | 511 | 2 | US-08-537- | Sequence 34, Applicati | 3.82e+01 |
| 19 | 56 | 53.3 | 515 | 2 | US-08-537- | Sequence 3, Applicatio | 3.82e+01 |
| 20 | 56 | 53.3 | 713 | 1 | US-08-332- | Sequence 56, Applicati | 3.82e+01 |
| 21 | 56 | 53.3 | 713 | 1 | US-08-332- | Sequence 62, Applicati | 3.82e+01 |
| 22 | 56 | 53.3 | 713 | 1 | US-08-188- | Sequence 62, Applicati | 3.82e+01 |
| 23 | 55 | 52.4 | 878 | 1 | US-08-237- | Sequence 2, Applicatio | 4.81e+01 |

| | | | | | | | |
|----|----|------|------|---|------------|------------------------|----------|
| 24 | 55 | 52.4 | 878 | 3 | PCT-US95-0 | Sequence 2, Applicatio | 4.81e+01 |
| 25 | 55 | 52.4 | 884 | 2 | US-08-474- | Sequence 8, Applicatio | 4.81e+01 |
| 26 | 55 | 52.4 | 884 | 2 | US-08-474- | Sequence 8, Applicatio | 4.81e+01 |
| 27 | 55 | 52.4 | 884 | 2 | US-08-472- | Sequence 7, Applicatio | 4.81e+01 |
| 28 | 54 | 51.4 | 246 | 2 | US-08-415- | Sequence 31, Applicati | 6.04e+01 |
| 29 | 53 | 50.5 | 199 | 2 | US-08-211- | Sequence 6, Applicatio | 7.57e+01 |
| 30 | 53 | 50.5 | 794 | 1 | US-08-332- | Sequence 60, Applicati | 7.57e+01 |
| 31 | 53 | 50.5 | 794 | 1 | US-08-332- | Sequence 54, Applicati | 7.57e+01 |
| 32 | 53 | 50.5 | 794 | 1 | US-08-188- | Sequence 60, Applicati | 7.57e+01 |
| 33 | 53 | 50.5 | 2647 | 2 | US-08-583- | Sequence 8, Applicatio | 7.57e+01 |
| 34 | 53 | 50.5 | 2647 | 2 | US-08-779- | Sequence 8, Applicatio | 7.57e+01 |
| 35 | 52 | 49.5 | 535 | 2 | US-08-007- | Sequence 2, Applicatio | 9.48e+01 |
| 36 | 52 | 49.5 | 1026 | 1 | US-08-268- | Sequence 95, Applicati | 9.48e+01 |
| 37 | 52 | 49.5 | 1026 | 1 | US-08-453- | Sequence 95, Applicati | 9.48e+01 |
| 38 | 52 | 49.5 | 1026 | 2 | US-08-453- | Sequence 95, Applicati | 9.48e+01 |
| 39 | 52 | 49.5 | 1026 | 1 | US-07-998- | Sequence 95, Applicati | 9.48e+01 |
| 40 | 52 | 49.5 | 1026 | 1 | US-08-453- | Sequence 95, Applicati | 9.48e+01 |
| 41 | 52 | 49.5 | 1026 | 3 | PCT-US93-1 | Sequence 95, Applicati | 9.48e+01 |
| 42 | 52 | 49.5 | 1203 | 1 | US-07-998- | Sequence 95, Applicati | 9.48e+01 |
| 43 | 52 | 49.5 | 1203 | 1 | US-08-268- | Sequence 103, Applicat | 9.48e+01 |
| 44 | 52 | 49.5 | 1203 | 3 | PCT-US93-1 | Sequence 103, Applicat | 9.48e+01 |
| 45 | 52 | 49.5 | 1203 | 1 | US-08-453- | Sequence 103, Applicat | 9.48e+01 |

ALIGNMENTS

RESULT 1
ID US-08-400-796-2 STANDARD: PRT: 15 AA.

XX xxxxxx

Sequence 2, Application US/08400796

Sequence 2, Application US/08400796

Patent No. 5874531

GENERAL INFORMATION:

APPLICANT: STROMINGER, JACK L.

TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

STREET: 600 ATLANTIC AVENUE

CITY: BOSTON

STATE: MA

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/400,796

FILING DATE:

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: TWOMEY, MICHAEL J.

REGISTRATION NUMBER: 38,349

REFERENCE/DOCKET NUMBER: H0498/7015

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

CC FRAGMENT TYPE: Internal
CC ORIGINAL SOURCE:
CC ORGANISM: HOMO SAPIENS
SQ SEQUENCE 15 AA: 1652 MW: 1115 CN:

Query Match 100.0% Score 105; DB 2; Length 15;
Best Local Similarity 100.0% Pred. No. 2.01e-04;
Matches 15: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FGIFVVDKNTGDIINI 15
OY 1 FGIFVVDKNTGDIINI 15

RESULT 2
ID US-08-738-349-12 STANDARD: PRT: 615 AA.
XX
XX
XX xxxxxx

DE Sequence 12, Application US/08738349

CC Sequence 12, Application US/08738349
CC Patent No. 5869638
CC GENERAL INFORMATION:

CC APPLICANT: Takeshita, Sunao
CC APPLICANT: Okazaki, Makoto
CC APPLICANT: Kawai, Shinji
CC APPLICANT: Tsujimura, Atsushi
CC APPLICANT: Amanu, Egon
CC TITLE OF INVENTION: Bone-Related Cadherin-Like Protein and
CC TITLE OF INVENTION: Process for Its Production
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC STREET: 1300 I Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20005-3315

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/738,349
CC FILING DATE: 25-OCT-1996
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/364,439
CC FILING DATE:
CC APPLICATION NUMBER: US 08/112,061
CC FILING DATE: 26-AUG-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Barker, M. P.
CC REGISTRATION NUMBER: 32,013
CC REFERENCE/DOCKET NUMBER: 02481.1323-00000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-408-4000
CC TELEFAX: 202-408-4000
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 615 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear

CC MOLECULE TYPE: Peptide
CC SEQUENCE 615 AA: 67754 MW: 1891114 CN;

Query Match 54.3% Score 57; DB 2; Length 615;
Best Local Similarity 50.0% Pred. No. 3.04e-01;

Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 94 IFVYIDKSGNII 105
OY 3 IFVVDKNTGDIIN 14

RESULT 3
ID US-08-738-349-6 STANDARD: PRT: 693 AA.
XX
XX
XX xxxxxx

DE Sequence 6, Application US/08738349

CC Sequence 6, Application US/08738349
CC Patent No. 5869638
CC GENERAL INFORMATION:

CC APPLICANT: Takeshita, Sunao
CC APPLICANT: Okazaki, Makoto
CC APPLICANT: Kawai, Shinji
CC APPLICANT: Tsujimura, Atsushi
CC APPLICANT: Amanu, Egon
CC TITLE OF INVENTION: Bone-Related Cadherin-Like Protein and
CC TITLE OF INVENTION: Process for Its Production
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC STREET: 1300 I Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20005-3315

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/738,349
CC FILING DATE: 25-OCT-1996
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/364,439
CC FILING DATE:
CC APPLICATION NUMBER: US 08/112,061
CC FILING DATE: 26-AUG-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Barker, M. P.
CC REGISTRATION NUMBER: 32,013
CC REFERENCE/DOCKET NUMBER: 02481.1323-00000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-408-4000
CC TELEFAX: 202-408-4000
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 693 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear

CC MOLECULE TYPE: Protein
CC SEQUENCE 693 AA: 76422 MW: 2487235 CN;

Query Match 54.3% Score 57; DB 2; Length 693;
Best Local Similarity 50.0% Pred. No. 3.04e-01;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 100 IFVYIDKSGNII 111
OY 3 IFVVDKNTGDIIN 14

```

DE      Sequence 58, Application US/08188228
XX
CC      Sequence 59, Application US/08188228
CC      Patent No. 5597725
CC      GENERAL INFORMATION:
CC      APPLICANT: Suzuki, Shintaro
CC      TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
CC      NUMBER OF SEQUENCES: 62
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
CC      ADDRESSEE: Borum
CC      STREET: 6300 Sears Tower, 233 S. Wacker Drive
CC      CITY: Chicago
CC      STATE: Illinois
CC      COUNTRY: USA
CC      ZIP: 60606
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Patentin Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/188,228
CC      FILING DATE:
CC      CLASSIFICATION: 435
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 08/049,460
CC      FILING DATE: 19 APR 1993
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 07/872,643
CC      FILING DATE: 17 APR 1992
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: No. 5597725and, Greta E.
CC      REGISTRATION NUMBER: 35,302
CC      REFERENCE/DOCKET NUMBER: 31340
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (312) 474-6300
CC      TELEFAX: (312) 474-0448
CC      TELEX: 25-3856
CC      INFORMATION FOR SEQ ID NO: 58:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 796 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 796 AA; 88049 MW; 3168744 CN;
SQ
Query Match          54.3%; Score 57; DB 1; Length 796;
Best Local Similarity 50.0%; Pctd. No. 3.04e+01;
Matches      6; Conservative      5; Mismatches      1; Indels      0; Gaps      0;

Db      100 IFVIDKSGNTH 111
      |||:|::|:|
QY      3 IFVVDKNTGTDIN 14

RESULT      6
ID      US-08-738-349-4      STANDARD:      PRT:      796 AA.
XX      xxxxxx
XX
XX      Sequence 4, Application US/08738349
CC      Sequence 4, Application US/08738349
CC      Patent No. 5869638
CC      GENERAL INFORMATION:
CC      APPLICANT: Takeshita, Sunao
CC      APPLICANT: Okazaki, Makoto
CC      APPLICANT: Kawai, Shinji
CC      APPLICANT: Tsujimura, Atsushi
CC      APPLICANT: Amann, Egon

```

CC TITLE OF INVENTION: Bone-Related Cadherin-Like Protein and
CC TITLE OF INVENTION: Process for Its Production
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC STREET: 1300 I Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20005-3315
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/738,349
CC FILING DATE: 25-OCT-1996
CC CLASSIFICATION: 435
CC PRIORITY APPLICATION DATA:
CC APPLICATION NUMBER: US 08/364,439
CC FILING DATE:
CC APPLICATION NUMBER: US 08/112,061
CC FILING DATE: 26-AUG-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Barker, M. P.
CC REGISTRATION NUMBER: 32,013
CC REFERENCE/DOCKET NUMBER: 02481.1323-00000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-408-4000
CC TELEFAX: 202-408-4400
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 796 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
CC SEQUENCE: 796 AA; 87931 MW; 3173092 CN;
DB 100 IFVYDKNSGNIH 111
|||:|:|:|:
OY 3 IFVVDKNTGDIIN 14
Query Match 54.3%; Score 57; DB 2; Length 796;
Best Local Similarity 50.0%; Pred. No. 3,04e+01;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
DE 100 IFVYDKNSGNIH 111
|||:|:|:|:
OY 3 IFVVDKNTGDIIN 14
RESULT 7
ID US-08-332-643-52 STANDARD; PRT: 796 AA.
XX xxxxxx
AC
DT
XX
XX
DE Sequence 52, Application US/08332643
XX Patent No. 5639634
CC GENERAL INFORMATION:
CC APPLICANT: Suzuki, Shintaro
CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
CC NUMBER OF SEQUENCES: 56
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
CC STREET: Bicknell, O'Toole, Gerstein, Murray &
CC STREET: Two First National Plaza, 20 South Clark
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60603
CC

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/332,643
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIORITY APPLICATION DATA:
CC APPLICATION NUMBER: US/07/872,643
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: No. 5639634and, Greta E.
CC REGISTRATION NUMBER: 35,302
CC REFERENCE/DOCKET NUMBER: 27866/30795
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (312) 346-5750
CC TELEFAX: (312) 984-9740
CC TELEX: 25-3856
CC INFORMATION FOR SEQ ID NO: 52:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 796 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
CC SEQUENCE: 796 AA; 88049 MW; 3168744 CN;
DB 100 IFVYDKNSGNIH 111
|||:|:|:|:
OY 3 IFVVDKNTGDIIN 14
Query Match 54.3%; Score 57; DB 1; Length 796;
Best Local Similarity 50.0%; Pred. No. 3,04e+01;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
DE 100 IFVYDKNSGNIH 111
|||:|:|:|:
OY 3 IFVVDKNTGDIIN 14
RESULT 8
ID US-08-332-638-58 STANDARD; PRT: 796 AA.
XX xxxxxx
AC
DT
XX
XX
DE Sequence 58, Application US/08332638
XX Patent No. 5646250
CC GENERAL INFORMATION:
CC APPLICANT: Suzuki, Shintaro
CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
CC NUMBER OF SEQUENCES: 62
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
CC STREET: 6300 Sears Tower, 233 S. Wacker Drive
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60606
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/332,638
CC FILING DATE: 01-NOV-1994
CC CLASSIFICATION: 435
CC PRIORITY APPLICATION DATA:
CC APPLICATION NUMBER: US 07/872,643
CC FILING DATE: 17 APR 1992
CC APPLICATION NUMBER: US/08/049,460
CC

CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: No. 5645250and, Greta E.
CC REGISTRATION NUMBER: 35,302
CC REFERENCE/DOCKET NUMBER: 31340
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (312) 474-6300
CC TELEFAX: (312) 474-0448
CC TELEX: 25-3856
CC INFORMATION FOR SEQ ID NO: 58:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 796 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 796 AA; 88049 MW; 3168744 CN;
SQ

Query Match 54.3%; Score 57; DB 1; Length 796;
Best Local Similarity 50.0%; Pred. No. 3.04e+01;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 100 IFVIDKSGNIH 111
| | | : : : |
3 IFVVDKNTGDI 14

OY

RESULT 9
ID US-08-268-161A-112 STANDARD; PRT: 797 AA.
XX xxxxxx
DT

Sequence 112, Application US/08268161A
CC Patent No. 5798224
CC GENERAL INFORMATION:
CC APPLICANT: Suzuki, Shintaro
CC TITLE OF INVENTION: Protocadherin Materials and Methods
CC NUMBER OF SEQUENCES: 115
CC CORRESPONDENCE ADDRESS:
CC ADDRESS: Marshall, O'Toole, Gerstein, Murray, &
CC ADDRESSEE: Borun
CC STREET: 233 South Wacker, 6300 Sears Tower
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60606
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/268.161A
CC FILING DATE: June 27, 1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Young J. Suh
CC REGISTRATION NUMBER: P-41,337
CC REFERENCE/DOCKET NUMBER: 27866/32149
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 312/474-6300
CC TELEFAX: 312/474-0448
CC TELEX: 25-3856
CC INFORMATION FOR SEQ ID NO: 112:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 797 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 797 AA; 87474 MW; 3358246 CN;
SQ

Query Match 54.3%; Score 57; DB 1; Length 797;
Best Local Similarity 46.7%; Pred. No. 3.04e+01;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

DB 288 FOFVIDETGEIHL 302
| | | : : : |
1 FGFVVDKNTGDI 15

OY

RESULT 10
ID US-08-453-695A-112 STANDARD; PRT: 797 AA.
XX xxxxxx
DT

Sequence 112, Application US/08453695A
CC Patent No. 5708143
CC GENERAL INFORMATION:
CC APPLICANT: Suzuki, Shintaro
CC TITLE OF INVENTION: Protocadherin Materials and Methods
CC NUMBER OF SEQUENCES: 115
CC CORRESPONDENCE ADDRESS:
CC ADDRESS: Marshall, O'Toole, Gerstein, Murray, &
CC ADDRESSEE: Borun
CC STREET: 233 South Wacker, 6300 Sears Tower
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60606
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/453.695A
CC FILING DATE:
CC CLASSIFICATION: 530
CC ATTORNEY/AGENT INFORMATION:
CC NAME: No. 570813and, Greta E.
CC REGISTRATION NUMBER: 35,302
CC REFERENCE/DOCKET NUMBER: 32658
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 312/474-6300
CC TELEFAX: 312/474-0448
CC TELEX: 25-3856
CC INFORMATION FOR SEQ ID NO: 112:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 797 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 797 AA; 87474 MW; 3358246 CN;
SQ

Query Match 54.3%; Score 57; DB 1; Length 797;
Best Local Similarity 46.7%; Pred. No. 3.04e+01;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

DB 288 FOFVIDETGEIHL 302
| | | : : : |
1 FGFVVDKNTGDI 15

OY

RESULT 11
ID US-08-453-702A-112 STANDARD; PRT: 797 AA.
XX xxxxxx
DT

XX Sequence 112, Application US/08453702A
 DE
 CC Sequence 112, Application US/08453702A
 CC Patent No. 5891706
 CC GENERAL INFORMATION:
 CC APPLICANT: Suzuki, Shintaro
 CC TITLE OF INVENTION: Protocadherin Materials and Methods
 CC NUMBER OF SEQUENCES: 115
 CC CORRESPONDENCE ADDRESS: 115
 CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
 CC STREET: 233 South Wacker, 6300 Sears Tower
 CC CITY: Chicago
 CC STATE: Illinois
 CC COUNTRY: USA
 CC ZIP: 60606
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patentin Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/453,702A
 CC FILING DATE:
 CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: No. 5891706and, Greta E.
 CC REGISTRATION NUMBER: 35,302
 CC TELEPHONE: 312/474-6300
 CC TELEFAX: 312/474-0448
 CC TELEX: 25-3856
 CC INFORMATION FOR SEQ ID NO: 112:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 797 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: Protein
 CC SEQUENCE 797 AA: 87474 MW: 3358246 CN:
 SO
 Query Match 54.3%; Score 57; DB 2; Length 797;
 Best Local Similarity 46.7%; Pred. No. 3.04e+01;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 Db 288 FGFVVDKNTGDI 302
 QY 1 FGFVVDKNTGDI 15
 RESULT 12 STANDARD: PRT: 797 AA.
 ID PCV-0595-08071-112
 AC xxxxxx
 XX
 DE Sequence 112, Application PC/TUS9508071
 CC GENERAL INFORMATION:
 CC APPLICANT: Suzuki, Shintaro
 CC TITLE OF INVENTION: Protocadherin Materials and Methods
 CC NUMBER OF SEQUENCES: 115
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
 CC STREET: 6300 Sears Tower, 233 S. Wacker Drive
 CC CITY: Chicago
 CC STATE: Illinois
 CC COUNTRY: USA
 CC ZIP: 60606
 CC

CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patentin Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: PCT/US95/08071
 CC FILING DATE:
 CC CLASSIFICATION:
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: PCT/US93/12588
 CC FILING DATE: 23 DEC 1993
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 07/998,003
 CC FILING DATE: 29 DEC 1992
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Noland, Greta E.
 CC REGISTRATION NUMBER: 35,302
 CC TELEPHONE: 312/474-6300
 CC TELEFAX: 312/474-0448
 CC TELEX: 25-3856
 CC INFORMATION FOR SEQ ID NO: 112:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 797 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: Protein
 CC SEQUENCE 797 AA: 87474 MW: 3358246 CN:
 SO
 Query Match 54.3%; Score 57; DB 3; Length 797;
 Best Local Similarity 46.7%; Pred. No. 3.04e+01;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 Db 288 FGFVVDKNTGDI 302
 QY 1 FGFVVDKNTGDI 15
 RESULT 13 STANDARD: PRT: 837 AA.
 ID US-08-472-481-6
 AC xxxxxx
 XX
 DE Sequence 6, Application US/08472481
 CC Patent No. 5863804
 CC GENERAL INFORMATION:
 CC APPLICANT: Ranscht, Barbara
 CC TITLE OF INVENTION: T-Cadherin Adhesion Molecule
 CC NUMBER OF SEQUENCES: 8
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Campbell and Flores
 CC STREET: 4370 La Jolla Village Drive, Suite 700
 CC CITY: San Diego
 CC STATE: California
 CC COUNTRY: United States
 CC ZIP: 92122
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patentin Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/472,481
 CC FILING DATE: 07-JUN-1995
 CC CLASSIFICATION: 424
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 08/213,361
 CC

CC FILING DATE: 14-MAY-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/607,293
CC FILING DATE: 30-OCT-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Campbell, Cathryn A.
CC REGISTRATION NUMBER: 31,815
CC REFERENCE/DOCKET NUMBER: P-LJ 1686
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 535-9001
CC TELEFAX: (619) 535-8949
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 837 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
SQ SEQUENCE 837 AA: 92281 MW: 3540645 CN:

Db 159 GIFFIERETG 168
|||:::|
OY 2 GIFFVDKNTG 11

Query Match 54.3%; Score 57; DB 2; Length 837;
Best Local Similarity 50.0%; Pred. No. 3.04e+01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DE Sequence 7, Application US/08474067
CC Sequence 7, Application US/08474067
CC Patent No. 581518
CC GENERAL INFORMATION:
CC APPLICANT: Ranscht, Barbara
CC TITLE OF INVENTION: T-Cadherin Adhesion Molecule
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Campbell and Flores
CC STREET: 4370 La Jolla Village Drive, Suite 700
CC CITY: San Diego
CC STATE: California
CC COUNTRY: United States
CC ZIP: 92122
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/474,067
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/213,361
CC FILING DATE: 14-MAY-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/607,293
CC FILING DATE: 30-OCT-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Campbell, Cathryn A.
CC REGISTRATION NUMBER: 31,815
CC REFERENCE/DOCKET NUMBER: P-LJ 1682
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 535-9001
CC TELEFAX: (619) 535-8949
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 837 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
SQ SEQUENCE 837 AA: 92281 MW: 3540645 CN:

CC LENGTH: 837 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
SQ SEQUENCE 837 AA: 92281 MW: 3540645 CN:

Db 159 GIFFIERETG 168
|||:::|
OY 2 GIFFVDKNTG 11

Query Match 54.3%; Score 57; DB 2; Length 837;
Best Local Similarity 50.0%; Pred. No. 3.04e+01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DE Sequence 7, Application US/08474068A
CC Sequence 7, Application US/08474068A
CC Patent No. 5837525
CC GENERAL INFORMATION:
CC APPLICANT: Ranscht, Barbara
CC TITLE OF INVENTION: T-Cadherin Adhesion Molecule
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Campbell & Flores LLP
CC STREET: 4370 La Jolla Village Drive, Suite 700
CC CITY: San Diego
CC STATE: California
CC COUNTRY: United States
CC ZIP: 92122
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/474,068A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/213,361
CC FILING DATE: 14-MAY-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/607,293
CC FILING DATE: 30-OCT-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Campbell, Cathryn A.
CC REGISTRATION NUMBER: 31,815
CC REFERENCE/DOCKET NUMBER: P-LJ 1683
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 535-9001
CC TELEFAX: (619) 535-8949
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 837 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
SQ SEQUENCE 837 AA: 92281 MW: 3540645 CN:

Db 159 GIFFIERETG 168
|||:::|
OY 2 GIFFVDKNTG 11

Query Match 54.3%; Score 57; DB 2; Length 837;
Best Local Similarity 50.0%; Pred. No. 3.04e+01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Mon May 15 08:08:03 2000

US-08-991-628-2a.ra1

Page 8

Search completed: Sat May 13 06:07:19 2000
Job time : 63 secs.

THIS PAGE BLANK (USPTO)

MPSRCH_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:10:47 2000; Maspar time 3.01 Seconds
Tabular output not generated. 117,892 Million cell updates/sec

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

(TM)

Title: >US-08-991-628-3
Description: (1-15) from US08991628.pep
Perfect Score: 96
Sequence: 1 LNSKIAFKIVSOEPA 15
Scoring table: PAM 150
Gap 15
Searched: 188963 seqs, 23686106 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: a.geneseq35
1:geneseqp
Statistics: Mean 18.170; Variance 51.971; scale 0.350

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|------------|-------|-------------|--------|-------|------------------------|-----------|
| 1 | 96 | 100.0 | 15 | 1 | Self epitope of desmog | 3.93e-04 |
| 2 | 96 | 100.0 | 15 | 1 | Desmoglein-3 190-204 | 3.93e-04 |
| 3 | 96 | 100.0 | 15 | 1 | Desmoglein 3 protein f | 3.93e-04 |
| 4 | 96 | 100.0 | 15 | 1 | Pemphigus vulgaris ant | 3.93e-04 |
| 5 | 96 | 100.0 | 15 | 1 | Human pemphigus vulgar | 3.93e-04 |
| 6 | 87 | 90.6 | 778 | 1 | Pemphigus foliaceus an | 5.50e-03 |
| 7 | 77 | 80.2 | 263 | 1 | Segment of desmosomal | 9.68e-02 |
| 8 | 77 | 80.2 | 560 | 1 | Sequence of desmosomal | 4.96e+01 |
| 9 | 54 | 56.3 | 878 | 1 | Human E-cadherin precu | 4.96e+01 |
| 10 | 54 | 56.3 | 2233 | 1 | Human E-cadherin wild | 4.96e+01 |
| 11 | 54 | 56.3 | 2233 | 1 | HPV-3 JS isolate wild | 4.96e+01 |
| 12 | 54 | 56.3 | 2233 | 1 | HPV-3 Vero CP45 vacci | 4.96e+01 |
| 13 | 54 | 56.3 | 2233 | 1 | HPV-3 F81 CP45 vacci | 4.96e+01 |
| 14 | 53 | 55.2 | 456 | 1 | Human cerebral protein | 6.41e+01 |
| 15 | 49 | 51.0 | 572 | 1 | Bilirubin oxidase | 1.75e+02 |
| 16 | 49 | 51.0 | 916 | 1 | Full length human cadh | 1.75e+02 |
| 17 | 49 | 51.0 | 916 | 1 | Human cadherin-4 | 1.75e+02 |
| 18 | 49 | 51.0 | 2408 | 1 | Translation of ORF 3 c | 1.75e+02 |
| 19 | 48 | 50.0 | 26 | 1 | Bacteriophage spo1 Pol | 2.25e+02 |
| 20 | 48 | 50.0 | 528 | 1 | H. pylori inner membra | 2.25e+02 |
| 21 | 48 | 50.0 | 2016 | 1 | Human hml sodium chann | 2.25e+02 |
| 22 | 48 | 50.0 | 2019 | 1 | Cardiac sodium channel | 2.25e+02 |
| 23 | 48 | 50.0 | 2020 | 1 | Cardiac sodium channel | 2.25e+02 |

| | | | | | | | |
|----|----|------|------|---|--------|-------------------------|----------|
| 24 | 47 | 49.0 | 198 | 1 | R06643 | Human growth hormone | 2.87e-02 |
| 25 | 47 | 49.0 | 308 | 1 | W75006 | Human secreted protein | 2.87e-02 |
| 26 | 47 | 49.0 | 330 | 1 | W89745 | Staphylococcus aureus | 2.87e-02 |
| 27 | 47 | 49.0 | 399 | 1 | W74883 | Human secreted protein | 2.87e-02 |
| 28 | 47 | 49.0 | 573 | 1 | R75741 | B31 outer surface prot | 2.87e-02 |
| 29 | 47 | 49.0 | 700 | 1 | R75731 | B. burgdorferi strain | 2.87e-02 |
| 30 | 47 | 49.0 | 716 | 1 | R30730 | B. burgdorferi 79 kd a | 2.87e-02 |
| 31 | 47 | 49.0 | 756 | 1 | R30177 | p1. | 2.87e-02 |
| 32 | 47 | 49.0 | 4472 | 1 | R87246 | Virulence gene cluster | 2.87e-02 |
| 33 | 46 | 47.9 | 33 | 1 | W79334 | Staphylococcus aureus | 3.66e-02 |
| 34 | 46 | 47.9 | 298 | 1 | R97629 | Human SLAMF2 T-cell co | 3.66e-02 |
| 35 | 46 | 47.9 | 305 | 1 | R97630 | Human SLAMF3 T-cell co | 3.66e-02 |
| 36 | 46 | 47.9 | 307 | 1 | R97631 | Human SLAMF4 T-cell co | 3.66e-02 |
| 37 | 46 | 47.9 | 335 | 1 | R97638 | Human SLAMF1 T-cell co | 3.66e-02 |
| 38 | 46 | 47.9 | 369 | 1 | Y07080 | Human cancer associate | 3.66e-02 |
| 39 | 46 | 47.9 | 459 | 1 | R13458 | P. denitrificans COB G. | 3.66e-02 |
| 40 | 46 | 47.9 | 481 | 1 | W34554 | MYTL glycosylase 29G. | 3.66e-02 |
| 41 | 46 | 47.9 | 486 | 1 | W38455 | Yeast RNA-binding prot | 3.66e-02 |
| 42 | 46 | 47.9 | 779 | 1 | R88226 | Rat neuronal protein k | 3.66e-02 |
| 43 | 46 | 47.9 | 821 | 1 | R99579 | Calpain large subunit | 3.66e-02 |
| 44 | 46 | 47.9 | 1275 | 1 | R13516 | P. denitrificans COB N. | 3.66e-02 |
| 45 | 46 | 47.9 | 1276 | 1 | R35199 | Mouse multidrug resist | 3.66e-02 |

ALIGNMENTS

| | | |
|--------|---|---------------------------|
| RESULT | 1 | standard: peptide: 15 AA. |
| ID | W04843 | |
| AC | W04843 | |
| DT | 18-FEB-1997 (first entry) | |
| DE | Self epitope of desmoglein 3, implicated in autoimmune disease. | |
| DM | Tolerisation: self-epitope; antigen: autoimmune disease. | |
| KM | autoantigen: HLA; human leukocyte antigen: T-cell; thymocyte; | |
| KW | pemphigus vulgaris; desmoglein; multiple sclerosis; | |
| KW | herpes simplex virus; adenovirus; phosphomannomutase; | |
| KW | human papillomavirus; Epstein-Barr virus; DNA polymerase; | |
| KW | influenza; haemagglutinin; reovirus; sigma protein. | |
| OS | Homo sapiens. | |
| PN | W09627387-A1. | |
| PD | 12-SEP-1996. | |
| PF | 07-MAR-1996: U03182. | |
| PR | 07-MAR-1995: US-400796. | |
| PI | (HARD) HARVARD COLLEGE. | |
| PA | Strominger JL, Mucherferfennly KW; | |
| DR | WPI: 96-425218/42. | |
| PT | Pemphigus vulgaris auto-antigens and multiple sclerosis non-self | |
| PT | antigens - useful in disease treatment, and method for | |
| PT | identification of other self and non-self antigens implicated in | |
| PT | auto-immune disease | |
| PS | Claim 1: Page 39; 58pp: English. | |
| CC | Pharmaceutical preparations for tolerisation to antigens comprise | |
| CC | either an isolated human non-collagen or non-mysin basic protein | |
| CC | (MBP) polypeptide which is capable of tolerising an individual to an | |
| CC | autoantigen; or an isolated human pathogen polypeptide capable of | |
| CC | tolerising an individual to that polypeptide. In both cases, the | |
| CC | polypeptide (whether self or non-self) includes an amino acid | |
| CC | sequence corresponding to a sequence motif for a MHC class II | |
| CC | protein, such as HLA-DR, which is associated with a human autoimmune | |
| CC | disease and which binds to the polypeptide to activate autoreactive | |
| CC | T-cells in individuals with the autoimmune disease. This peptide is | |
| CC | derived from the human desmoglein 3 protein (amino acids 190-204) | |
| CC | and is implicated as a self epitope in pemphigus vulgaris. Peptides | |
| CC | derived from the human desmoglein protein are described in W04841-47. | |
| SO | Sequence 15 AA: | |

Query Match 100.0%; Score 96; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.93e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 LNSKIAFKIVSOEPA 15
QY 1 LNSKIAFKIVSOEPA 15

RESULT 2
ID W64815 standard: peptide: 15 AA.
AC W64815:
DT 29-SEP-1998 (first entry)
DE Desmoglein-3 190-204.
KW Desmoglein: DG: gene therapy; pemphigus vulgaris; microparticle:
KW autoantigen; autoimmune disease; MHC.
OS Homo sapiens.
PN US5783567-A.
PD 21-JUL-1998.
PF 22-JAN-1997: 787547.
PR 22-JAN-1997: US-787547.
PA (PANG-) PANCAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS:
DR WPI: 98-427077/36.
PT Microparticle encapsulated nucleic acids - for recombinant
PT expression of proteins e.g. in gene therapy
PS Disclosure: Column 4: 42pp: English.
CC The patent describes a new preparation of microparticles each
CC comprising a polymeric matrix and a nucleic acid. The polymeric
CC matrix consists of one or more synthetic polymers having a solubility
CC in water of less than 1 mg/l (e.g. poly-lactic-co-glycolic acid);
CC and at least 90% of the microparticles have a diameter of less than
CC 100 microns. The microparticles are useful for the delivery of nucleic
CC acids to phagocytic cells. In one embodiment the microparticles are
CC less than 20 microns in diameter and the nucleic acid (preferably in
CC closed circular form) includes an expression control sequence
CC operatively linked to a coding sequence, where the expression product
CC of the coding sequence is a polypeptide having a length and a sequence
CC which permits it to bind to an MHC class I or II molecule. The
CC expression product is thus an effective stimulator of an immune
CC response in mammals. The present sequence, an antigenic portion of
CC desmoglein 3, is an example of an MHC class II peptide which can be
CC expressed by the nucleic acid. It is associated with pemphigus
CC vulgaris.
SQ Sequence 15 AA:
Query Match 100.0%; Score 96; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.93e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 LNSKIAFKIVSOEPA 15
Oy 1 LNSKIAFKIVSOEPA 15
RESULT 3
ID W78814 standard: peptide: 15 AA.
AC W78814:
DT 17-NOV-1998 (first entry)
DE Desmoglein 3 protein fragment 190-204.
KW Microparticle: delivery: polymeric matrix; autoantigen; tumour antigen;
KW class II associated peptide; pathogen; gene therapy; genetic disease;
KW infection; downregulation; immune response.
OS Homo sapiens.
PN Synthetic.
PN WO9831398-A1.
PD 23-JUL-1998.
PF 22-JAN-1998.
PR 06-JAN-1998: US-003253.
PR 22-JAN-1997: US-787547.
PA (PANG-) PANCAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS, Lunsford LB:
DR WPI: 98-427556/36.
PT New preparations of microparticles - comprising a synthetic polymer
PT matrix and nucleic acid comprising an expression vector for use in
PT gene therapy
PS Disclosure: Page 8: 101pp: English.
CC A microparticle preparation (MP) has been developed, consisting of
CC microparticles having a diameter of less than 100 mu m. The MP comprises:
CC (a) a polymeric matrix (PM) consisting of one or more synthetic polymers
CC having a solubility in water of less than 1 mg/l; and (b) an expression

CC vector selected from RNA molecules (at least 50% of which are closed
CC circles) or circular plasmid DNA (at least 50% of which are supercoiled).
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)
CC a PM; and (b) a NAM comprising an expression control sequence operatively
CC linked to a coding sequence, where the coding sequence encodes an
CC expression product selected from: (1) a polypeptide at least 7 amino
CC acids in length, having a sequence identical to the sequence of: (1) a
CC fragment of a naturally-occurring mammalian protein; or (11) a fragment
CC of a naturally-occurring protein from an infectious agent which infects
CC a mammal; (2) a peptide having a length and sequence which permits it to
CC bind to an MHC class I or II molecule; and (3) the polypeptide or the
CC peptide linked to a trafficking sequence. W69763 to W69765, and W7879?
CC to W78897 are peptide fragments for use in the present invention. The
CC MPs are highly effective vehicles for the delivery of polynucleotides
CC into phagocytic cells. They can be used for gene therapy, e.g. for
CC treating genetic diseases, infections or tumours or for downregulating
CC an immune response.
SQ Sequence 15 AA:
Query Match 100.0%; Score 96; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.93e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 LNSKIAFKIVSOEPA 15
Oy 1 LNSKIAFKIVSOEPA 15
RESULT 4
ID W07908 standard: protein: 614 AA.
AC W07908:
DT 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody: immunoglobulin G; IgG1; fusion protein; diagnosis;
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.
OS Homo sapiens.
PN J08188540-A.
PD 23-JUL-1996.
PF 30-JUN-1995: 165632.
PR 30-JUN-1994: JP-173291.
PA (NISH/) NISHIKAWA T.
DR WPI: 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris auto:antibody -
PT useful to treat and diagnose pemphigus vulgaris
PS Claim 1; Page 7-9; 9pp: Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
CC (vesicles) of the skin and mucous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose
CC pemphigus vulgaris.
SQ Sequence 614 AA:
Query Match 100.0%; Score 96; DB 1; Length 614;
Best Local Similarity 100.0%; Pred. No. 3.93e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 189 LNSKIAFKIVSOEPA 203
Oy 1 LNSKIAFKIVSOEPA 15
RESULT 5
ID R30742 standard: Protein: 999 AA.
AC R30742:
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130KD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.

PN US798918-A.
PD 15-DEC-1992.
PR 27-NOV-1991: 798918.
PI (US98) US DEPT HEALTH & HUMAN SERVICE.
PA Amagai M, Klaus-Kovtun V, Stanley JR;
DR WPI: 93-067436/08.
DR N-PSDB: 035992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses
PS Disclosure: Fig 7: 50pp: English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA:

Query Match 100.0%; Score 96; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 3,936-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 190 LNSKIAKIVSOEPA 204
OY 1 LNSKIAKIVSOEPA 15
|||||

RESULT 6
ID W15489 standard; protein: 778 AA.
AC W15489;
DT 17-JUN-1997 (first entry)
DE Pemphigus foliaceus antigen-IgG constant region fusion protein.
KW Pemphigus foliaceus; autoantibody; constant region; IgG;
KW extracellular region; antigen; hinge portion; skin;
KW dermatitis herpetiformis; fusion protein; detection; ss.
OS Chimeric - Homo sapiens.
FH Key Location/Qualifiers
FT domain 1..545
FT /note= "Pemphigus foliaceus antigen protein"
FN J09077800-A.
PD 25-MAR-1997.
PR 12-SEP-1995: 260899.
PR 12-SEP-1995: JP-260899.
PA (NISH/) NISHIKAWA T.
DR WPI: 97-241758/22.
DR P-PSDB: T66428.
PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
PT through the hinge region used to treat pemphigus foliaceus
PS Claim 1: Page 10-12: 17pp: Japanese.
CC This sequence represents a fused protein recognised by pemphigus
CC foliaceus patient autoantibody which comprises the constant region
CC of IgG linked to the extracellular region of pemphigus foliaceus
CC antigen protein through the hinge portion. Pemphigus foliaceus is
CC a chronic, generalised, vesicular and scaling skin eruption similar
CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion
CC protein is useful to treat pemphigus foliaceus. The antigen is
CC especially administered through an adsorbent upon which the fusion
CC protein is immobilised via a carrier. The fusion protein is also
CC useful for detecting pemphigus foliaceus antibodies which is useful
CC in immunodiagnosis. The fusion protein has little or no side effects.
SQ Sequence 778 AA:

Query Match 90.6%; Score 87; DB 1; Length 778;
Best Local Similarity 80.0%; Pred. No. 5,506-03;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 LNSKIAKIVSOEPA 204
OY 1 LNSKIAKIVSOEPA 15
|||||

RESULT 7
ID W13010 standard; protein: 263 AA.
AC W13010;

DT 21-NOV-1997 (first entry)
DE Segment of desmosomal cadherin, desmoglein Dsg2.
KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW micrometastasis; separation; enrichment; targeted delivery;
KW metastatic.
OS Homo sapiens.
PN DE19531033-A1.
PD 27-FEB-1997.
PR 23-AUG-1995: 031033.
PR 23-AUG-1995: DE-031033.
PA (PROG-) PROGEN BIOTECHNIK GMBH.
PI Franke WW, Schaefer S;
DR WPI: 97-146518/14.
PT Antibody reactive with part of desmosomal cadherin - exposed on
PT surface of epithelial or carcinoma cells, not bound to desmosomes,
PT useful for diagnosis and treatment of carcinoma micrometastases
PS Claim 9: Page 5; 8pp: German.
CC The present sequence is a segment of the desmosomal cadherin (DC),
CC desmoglein Dsg2, which is exposed on the surface of epithelial or
CC carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC directed against epitopes of the present sequence can be used to
CC diagnose, i.e. to detect carcinoma cells, especially
CC micrometastases, not bound to desmosomes, to separate, enrich or
CC detect living or fixed carcinoma cells by cell sorting methods and
CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC target cells. The Ab provides rapid and reliable detection of
CC metastatic carcinoma, and detects parts of DC that are not
CC accessible in desmosome bound cells, as in normal tissue or
CC carcinoma.
SQ Sequence 263 AA:

Query Match 80.2%; Score 77; DB 1; Length 263;
Best Local Similarity 73.3%; Pred. No. 9,686-02;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 41 LNSKISRYVSLPEA 55
OY 1 LNSKIAKIVSOEPA 15
|||||

RESULT 8
ID W13009 standard; protein: 560 AA.
AC W13009;
DT 21-NOV-1997 (first entry)
DE Segment of desmosomal cadherin, desmoglein Dsg2.
KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW micrometastasis; separation; enrichment; targeted delivery;
KW metastatic.
OS Homo sapiens.
PN DE19531033-A1.
PD 27-FEB-1997.
PR 23-AUG-1995: 031033.
PR 23-AUG-1995: DE-031033.
PA (PROG-) PROGEN BIOTECHNIK GMBH.
PI Franke WW, Schaefer S;
DR WPI: 97-146518/14.
PT Antibody reactive with part of desmosomal cadherin - exposed on
PT surface of epithelial or carcinoma cells, not bound to desmosomes,
PT useful for diagnosis and treatment of carcinoma micrometastases
PS Claim 7: Page 5; 8pp: German.
CC The present sequence is a segment of the desmosomal cadherin (DC),
CC desmoglein Dsg2, which is exposed on the surface of epithelial or
CC carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC directed against epitopes of the present sequence can be used to
CC diagnose, i.e. to detect carcinoma cells, especially
CC micrometastases, not bound to desmosomes, to separate, enrich or
CC detect living or fixed carcinoma cells by cell sorting methods and
CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC target cells. The Ab provides rapid and reliable detection of
CC metastatic carcinoma, and detects parts of DC that are not
CC accessible in desmosome bound cells, as in normal tissue or

CC Carcinomas. 560 AA.
S0 Sequence 560 AA.
Query Match 80.2%; Score 77; DB 1; Length 560;
Best Local Similarity 73.3%; Pred. No. 9,68e+02;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
DB 143 LNSKISRYVLSLEPA 157
OY 1 LNSKIAFKIVSOEPA 15
RESULT 9
ID R55060 standard; Protein: 878 AA.
AC R55060:1994 (first entry)
DE Sequence of human liver E-cadherin.
KW Human epithelial-cadherin; E-cadherin; cell adhesion molecule; CAM;
KM uvomorulin; L-CAM; Cell CAM 120/80.
OS Homo sapiens.
PN M09411401-A.
PD 26-MAY-1994.
PF 16-NOV-1993: U11097.
PR 17-NOV-1992: US-978897.
PA (UYVA) UNIV YALE.
PI Morrow JS, Rimm DL;
DR MPI: 94-183426/22.
N-PSDB: 065487.
PT Purified human E-cadherin protein and nucleic acid - used to
develop prods. for diagnosis, prognosis, therapy and prophylaxis
of E-cadherin disorders, e.g. malignancies
PS Claim 1: Page 59-63; 97pp; English.
CC E-cadherin is a cell adhesion molecule that is also known as
uvomorulin. L-CAM and Cell CAM 120/80. The DNA encoding hEC was obtd.
CC by screening normal human liver and hepatocellular carcinoma cDNA
libraries and a colonic epithelial cell cDNA library. The following
sequences are specifically claimed: AAs 1-878; 151-878; 30
CC 401; AAs 402-513; AAs 178-878; AAs 1-150; AAs 178-289; AAs 290-
CC AAs 704-878; nucleotide sequences comprising nucleotide numbers 116-
CC 2749; 566-2749; 1-1053; 510-2696; 1332-3000; 540-1500; 348-906; 890-
CC 1648; 384-1208; 641-2046; 685-1336; 880-1661; 1199-1742; 1373-1742;
CC 1705-2204; 2458-2775; DNA encoding at least 30 AAs selected from
CC AAs 308-878. The prods. can be used in the diagnosis, prognosis,
CC therapy and prophylaxis of conditions involving improper E-cadherin
expression. Suitable dosages for i.v. admin. of a protein are
CC 20-500 mcg/kg body wt.
S0 Sequence 878 AA.
Query Match 56.3%; Score 54; DB 1; Length 878;
Best Local Similarity 53.8%; Pred. No. 4,96e+01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
DB 293 NAAIVYTLISODP 305
OY 2 NSKIAFKIVSOEP 14
RESULT 10
ID R85487 standard; Protein: 878 AA.
AC R85487:
DE 18-MAR-1996 (first entry)
KW Human E-cadherin precursor.
KM E-cadherin; T-lymphocyte; alpha-E; beta-7 integrin; cell adhesion;
KW autoimmune disease; Crohn disease; psoriasis.
OS Homo sapiens.
PN Key
PD Location/Qualifiers
PF 1.150
FT peptide
FT 151..702
FT domain
FT /label= Sig-peptide
FT /label= Extracellular-domain
FT /note= "the extracellular domain (amino acids 1-552
FT of the mature protein) is the preferred

FT region for generation of peptides of the
FT invention.
FT 703..726
FT /label= Transmembrane-domain
FT 727..876
FT /label= Cytoplasmic-domain
PN M09529693-A1.
PD 09-NOV-1995.
PF 03-MAY-1995: U05518.
PR 03-MAY-1994: US-237919.
PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
PI Brenner MB, Ceppek KL;
DR MPI: 95-392921/50.
N-PSDB: T05764.
DR Inhibiting adhesion of T lymphocytes with E-cadherin - useful for
PT isolating agents to treat auto-immune diseases e.g. Crohn's disease,
PT psoriasis, etc
PS Disclosure: Page 70-75; 103pp; English.
CC The human E-cadherin protein precursor (R85487) is expressed by
CC an cDNA clone (T05764) derived from human liver. The extracellular
CC domain of E-cadherin is used to generate peptides that specifically
CC bind to heterotypic cognates of E-cadherin and which inhibit adhesion
CC of intra-epithelial T-lymphocytes to E-cadherin-expressing epithelial
CC or endothelial cells in vitro or in vivo, thereby modulating
CC mucosal immune responses. Such peptides are also specifically
CC reactive with a monoclonal antibody (E4.6 or E6.1) that binds to
CC E-cadherin and that can inhibit T-cell binding.
S0 Sequence 878 AA.
Query Match 56.3%; Score 54; DB 1; Length 878;
Best Local Similarity 53.8%; Pred. No. 4,96e+01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
DB 293 NAAIVYTLISODP 305
OY 2 NSKIAFKIVSOEP 14
RESULT 11
ID W48711 standard; Protein: 2233 AA.
AC W48711:
DE 13-OCT-1998 (first entry)
DE HPIV-3 JS isolate wild-type L protein.
KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KM single stranded RNA virus; Mononegavirales.
OS Human parainfluenza virus.
PN W09813501-A2.
PD 02-APR-1998.
PF 19-SEP-1997: U16718.
PR 27-SEP-1996: US-026823.
PA (AMCY) AMERICAN CYANAMID CO.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
DR MPI: 98-230710/20.
N-PSDB: V18272.
PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT single stranded RNA virus of order Mononegavirales - having
PT attenuating mutation in 3' genomic promoter region and RNA
PT polymerase gene, useful as vaccine to immunise against such virus
PS Disclosure: Page 246-254; 420pp; English.
CC This sequence represents the wild-type L protein from Human parainfluenza
CC virus (HPIV-3) type 3 isolate JS. This sequence is used in a method which
CC involves the isolation of recombinantly-generated, attenuated,
CC non-segmented, negative-sense, single stranded RNA virus of the order
CC Mononegavirales which have at least 1 attenuating mutation in the 3'
CC genomic promoter region and at least 1 attenuating mutation in the RNA
CC polymerase gene. This RNA virus can be used as a vaccine to immunise an
CC individual against such a virus.
S0 Sequence 2233 AA.
Query Match 56.3%; Score 54; DB 1; Length 2233;
Best Local Similarity 33.3%; Pred. No. 4,96e+01;
Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 973 LDRSVLYRINNOEPG 987
Y 1 LNSKIAFKIVSOEPA 15

RESULT 12
ID W48713 standard: Protein: 2233 AA.
AC W48713;
DT 13-OCT-1998 (first entry)
DE HPIV-3 Vero cp45 vaccine L protein.
KW L protein: attenuation; non-segmented; negative sense; vaccine; immunity;
KM single stranded RNA virus: Mononegavirales.
OS Human parainfluenza virus.
PN MO9813501-A2.
PD 02-APR-1998.
PF 19-SEP-1997: U16718.
PR 27-SEP-1996: US-026823.
PA (AMCY) AMERICAN CYANAMID CO.
PI (USSH) US DEPT HEALTH & HUMAN SERVICES.
DR Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
WPI: 98-230710/20.
DR N-PSDB: V18274.
PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT single stranded RNA virus of order Mononegavirales - RNA
PT attenuating mutation in 3' genomic promoter region and having
PT polymerase gene, useful as vaccine to immunise against such virus
PS Disclosure: Page 283-291: 426pp: English.
CC This sequence represents the Human parainfluenza virus (HPIV-3) type 3
CC vaccine Vero cp45 L protein. This sequence is used in a method which
CC involves the isolation of recombinantly-generated, attenuated,
CC non-segmented, negative-sense, single stranded RNA virus of the order
CC Mononegavirales which have at least 1 attenuating mutation in the 3'
CC genomic promoter region and at least 1 attenuating mutation in the RNA
CC polymerase gene. This RNA virus can be used as a vaccine to immunise an
CC individual against such a virus.
SQ Sequence 2233 AA;

Query Match 56.3%; Score 54; DB 1; Length 2233;
Best Local Similarity 33.3%; Pred. No. 4.96e+01;
Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 973 LDRSVLYRINNOEPG 987
Y 1 LNSKIAFKIVSOEPA 15

RESULT 13
ID W48712 standard: Protein: 2233 AA.
AC W48712;
DT 13-OCT-1998 (first entry)
DE HPIV-3 FRh1 cp45 vaccine L protein.
KW L protein: attenuation; non-segmented; negative sense; vaccine; immunity;
KM single stranded RNA virus: Mononegavirales.
OS Human parainfluenza virus.
PN MO9813501-A2.
PD 02-APR-1998.
PF 19-SEP-1997: U16718.
PR 27-SEP-1996: US-026823.
PA (AMCY) AMERICAN CYANAMID CO.
PI (USSH) US DEPT HEALTH & HUMAN SERVICES.
DR Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
WPI: 98-230710/20.
DR N-PSDB: V18273.
PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT single stranded RNA virus of order Mononegavirales - having
PT attenuating mutation in 3' genomic promoter region and RNA
PT polymerase gene, useful as vaccine to immunise against such virus
PS Disclosure: Page 265-273: 426pp: English.
CC This sequence represents the Human parainfluenza virus (HPIV-3) type 3
CC vaccine FRh1 cp45 L protein. This sequence is used in a method which
CC involves the isolation of recombinantly-generated, attenuated,
CC non-segmented, negative-sense, single stranded RNA virus of the order

CC Mononegavirales which have at least 1 attenuating mutation in the 3'
CC genomic promoter region and at least 1 attenuating mutation in the RNA
CC polymerase gene. This RNA virus can be used as a vaccine to immunise an
CC individual against such a virus.
SQ Sequence 2233 AA;

Query Match 56.3%; Score 54; DB 1; Length 2233;
Best Local Similarity 33.3%; Pred. No. 4.96e+01;
Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 973 LDRSVLYRINNOEPG 987
Y 1 LNSKIAFKIVSOEPA 15

RESULT 14
ID W82841 standard: Protein: 456 AA.
AC W82841;
DT 03-FEB-1999 (first entry)
DE Human cerebral protein-1
KW Human; cerebral protein-1; HUCEP-1; Ischaemic brain disease;
KW neuron function activating activity; nerve denatured disease;
KW Parkinson's disease; Alzheimer's disease.
OS Homo sapiens.
PN J10257891-A.
PD 29-SEP-1998.
PF 19-MAR-1997: 065716.
PR 19-MAR-1997: JP-065716.
PA (TAIS) TAISHO PHARM CO LTD.
DR WPI: 98-575902/49.
DR N-PSDB: V64162.
PT HUCEP-1 protein having neuron function activating activity - useful
PT for treating ischaemic brain diseases and nerve denaturation
PT conditions such as Parkinson's and Alzheimer's diseases
PS Calm 1; F1g 4; 17pp: Japanese.
CC The present sequence represents human cerebral protein-1 (HUCEP-1).
CC HUCEP-1 has neuron function activating activity. HUCEP-1 is useful
CC for treating ischaemic brain diseases and nerve denaturation
CC conditions such as Parkinson's and Alzheimer's diseases.
SQ Sequence 456 AA;

Query Match 55.2%; Score 53; DB 1; Length 456;
Best Local Similarity 54.5%; Pred. No. 6.41e+01;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 408 TKUGFKIVSKD 418
Y 3 SKIAFKIVSOE 13

RESULT 15
ID R40843 standard: Protein: 572 AA.
AC R40843;
DT 24-FEB-1994 (first entry)
DE Bilirubin oxidase.
KW Bilirubin; oxidase; analytical; BO; expression vector; PCR;
KM polymerase chain reaction.
OS Myrothecium verrucaria.
FH key Location/Qualifiers
FT peptide 1..38
FT /label= sig_peptide
FT protein 39..534
FT /label= mat_protein
PN J05199882-A.
PD 10-AUG-1993.
PF 24-JAN-1992: 034126.
PR 24-JAN-1992: JP-034126.
PA (AMAN) AMANO PHARM KK.
DR WPI: 93-284681/36.
DR N-PSDB: 047790.
PT Bilirubin oxidase prepn. useful as an analytical enzyme - by
PT culturing Bilirubin oxidase in transformant culture
PS Claim 1; Page 29-32; 32pp; Japanese.

Mon May 15 08:08:21 2000

US-08-991-628-3a.rag

Page 6

CC The sequence encodes biliverdin oxidase. The protein produced has a
CC 38 amino acid signal peptide which is removed to give the mature
CC protein (R40843).
CC
CC Sequence 572 AA;
CC

| | | | | |
|-----------------------|--------|---------------------|-------|-------------|
| Query Match | 51.0%; | Score 49; | DB 1; | Length 572; |
| Best Local Similarity | 13.3%; | Pred. No. 1.75e+02; | | |
| Matches | 2; | Conservative | 1; | Indels 0; |
| | | | | Gaps 0; |

```
Db      282 IDTRLPEKVIASDSG 296
        :::::|::: :::
QY      1 LNSKIAFKIVSQEPA 15
```

Search completed: Sat May 13 07:10:55 2000
Job time : 8 secs.

MUSE
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:10:21 2000; Maspar time 4.28 Seconds
Tabular output not generated. 165.192 Million cell updates/sec

Title: >US-08-991-628-3
Description: (1-15) from US08991628.pep
Perfect Score: 96
Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table:
PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: plr62
1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 25.750; Variance 32.305; scale 0.797

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|------------|-------|-------------|--------|-------|-------------------------------|-----------|
| 1 | 96 | 100.0 | 999 | 1 | IJHUG3 desmoglein 3 precursor | 1.15e-09 |
| 2 | 87 | 90.6 | 1043 | 1 | IJBOG1 desmoglein 1 precursor | 2.27e-07 |
| 3 | 87 | 90.6 | 1049 | 1 | IJHUG1 desmoglein 1 precursor | 2.27e-07 |
| 4 | 77 | 80.2 | 1117 | 2 | S38673 desmoglein 2 - human | 6.37e-05 |
| 5 | 61 | 63.5 | 1421 | 2 | T02501 hypothetical protein | 2.63e-01 |
| 6 | 60 | 62.5 | 884 | 1 | IJM5CE E-cadherin precursor | 4.27e-01 |
| 7 | 60 | 62.5 | 884 | 2 | S34438 uvomolulin - mouse | 4.27e-01 |
| 8 | 60 | 62.5 | 906 | 1 | IJHUCN cadherin 2 precursor | 4.27e-01 |
| 9 | 59 | 61.5 | 385 | 2 | S49752 homeotic protein YOX1 | 6.90e-01 |
| 10 | 58 | 60.4 | 138 | 2 | S74520 hypothetical protein | 1.11e+00 |
| 11 | 57 | 59.4 | 1264 | 2 | S64146 probable membrane pro | 1.77e+00 |
| 12 | 57 | 59.4 | 5147 | 1 | IJFFTH cadherin-related tumo | 4.45e+00 |
| 13 | 55 | 57.3 | 192 | 2 | F70126 translation elongatio | 7.00e+00 |
| 14 | 54 | 56.3 | 108 | 2 | T05921 N-cardanyl-L-amino ac | 7.00e+00 |
| 15 | 54 | 56.3 | 877 | 1 | IJBOCN N-cadherin precursor | 7.00e+00 |
| 16 | 54 | 56.3 | 882 | 1 | IJHUCE cadherin 1 precursor | 7.00e+00 |
| 17 | 54 | 56.3 | 905 | 2 | S43064 cadherin 1 African cl | 7.00e+00 |
| 18 | 54 | 56.3 | 906 | 1 | IJM5CN N-cadherin precursor | 7.00e+00 |
| 19 | 54 | 56.3 | 1146 | 2 | S64402 probable transcriptio | 7.00e+00 |
| 20 | 54 | 56.3 | 2233 | 1 | LNZNP3 genome polypeptide | 1.09e+01 |
| 21 | 53 | 55.2 | 775 | 2 | T07172 subtilisin-like prote | 1.09e+01 |
| 22 | 53 | 55.2 | 887 | 1 | IJCHCL E-cadherin precursor | 1.09e+01 |
| 23 | 53 | 55.2 | 2048 | 1 | LNZNSB genome polypeptide | 1.09e+01 |

| 24 | 53 | 55.2 | 2228 | 1 | LNZNSV genome polypeptide - | 1.09e+01 |
|----|----|------|------|---|------------------------------|----------|
| 25 | 52 | 54.2 | 191 | 2 | C72411 Holliday junction DNA | 1.70e+01 |
| 26 | 52 | 54.2 | 485 | 2 | A17266 probable transcriptio | 1.70e+01 |
| 27 | 52 | 54.2 | 732 | 1 | IJCHCB B-cadherin precursor | 1.70e+01 |
| 28 | 52 | 54.2 | 821 | 1 | B34488 calpain (EC 3.4.22.17 | 1.70e+01 |
| 29 | 51 | 53.1 | 241 | 2 | S63634 ribosomal protein S3 | 2.62e+01 |
| 30 | 51 | 53.1 | 293 | 2 | F49846 spa33 protein - Shige | 2.62e+01 |
| 31 | 51 | 53.1 | 293 | 2 | E42284 spa33 protein - Shige | 2.62e+01 |
| 32 | 51 | 53.1 | 644 | 2 | G64938 hypothetical protein | 2.62e+01 |
| 33 | 51 | 53.1 | 742 | 2 | A49341 isocitrate dehydrogen | 2.62e+01 |
| 34 | 51 | 53.1 | 813 | 3 | T02672 hypothetical protein | 2.62e+01 |
| 35 | 51 | 53.1 | 947 | 2 | D72067 polymorphic membrane | 2.62e+01 |
| 36 | 51 | 53.1 | 2764 | 2 | T13949 neurofibromin - fruit | 2.62e+01 |
| 37 | 51 | 53.1 | 2802 | 2 | T13945 neurofibromin - fruit | 2.62e+01 |
| 38 | 51 | 53.1 | 2802 | 2 | T13947 neurofibromin - fruit | 2.62e+01 |
| 39 | 50 | 52.1 | 217 | 2 | S25314 aspartic proteinase 1 | 4.01e+01 |
| 40 | 50 | 52.1 | 277 | 2 | JC5284 carbonyl reductase (N | 4.01e+01 |
| 41 | 50 | 52.1 | 355 | 2 | T15203 hypothetical protein | 4.01e+01 |
| 42 | 50 | 52.1 | 461 | 2 | A71662 sodium/pantothenate s | 4.01e+01 |
| 43 | 50 | 52.1 | 458 | 2 | A71662 hypothetical protein | 4.01e+01 |
| 44 | 50 | 52.1 | 578 | 2 | T02292 hypothetical protein | 4.01e+01 |
| 45 | 50 | 52.1 | 1045 | 2 | T13481 genome polypeptide 1 | 4.01e+01 |
| | | | 2410 | 1 | J01948 | |

ALIGNMENTS

| RESULT | ENTRY | 1 |
|-----------------|------------------------|---|
| ENTRY | IJHUG3 | #type complete |
| ALTERNATE_NAMES | desmoglein 3 precursor | - human |
| ORGANISM | pemphigus vulgaris | antigen |
| DATE | 30-Jun-1993 | #sequence_revision 30-Jun-1993 #text_change 23-Jun-1999 |

| ACCESSIONS | REFERENCE |
|------------|---|
| A41088 | Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R. |
| A41088 | Cell (1991) 67:869-877 |

| #authors | #journal | #title |
|----------|----------|---|
| 1 | Journal | Antibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion. |

| #cross-references | #label |
|-------------------|---|
| W6483 | NID:9190751; PID:AA60230.1; PID:9190751 |

| #cross-references | #label |
|-------------------|---|
| W6483 | NID:9190751; PID:AA60230.1; PID:9190751 |

GENETICS

| #gene | #cross-references |
|----------|-------------------------|
| GDB:DSG3 | GDB:134030; OMIM:169615 |

| #map_position | #superfamily |
|-----------------|--------------------------|
| 18q12.1-18q12.2 | cadherin repeat homology |

| #keywords | #binding |
|--|----------|
| calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein | |

| #feature | #domain |
|----------|---|
| 1-23 | domain signal sequence #status predicted #label SIG |
| 24-49 | domain propeptide #status predicted #label PRO |
| 50-999 | product desmoglein homolog #status predicted #label MAT |

| #domain | #status | #label |
|--|-------------------|--------|
| domain extracellular | predicted | EXT |
| domain cadherin repeat | homology | CR1 |
| domain cadherin repeat | homology | CR2 |
| domain cadherin repeat | homology | CR3 |
| domain cadherin repeat | homology | CR4 |
| domain cadherin repeat | homology | CR5 |
| domain transmembrane | #status predicted | TMN |
| domain intracellular | #status predicted | INT |
| domain desmoglein repeat | #label D51 | |
| domain desmoglein repeat | #label D52 | |
| binding-site carbohydrate (Asn) (covalent) | #status predicted | |

| #length | #molecular-weight | #checks |
|---------|-------------------|---------|
| 999 | 107502 | 8311 |

| Query Match | Score | DB 1 | Length |
|------------------------------|-------|----------|--------|
| Best Local Similarity 100.0% | 96 | 1.15e-09 | 999 |

| Matches | 15 | Conservative | 0 | Mismatches | 0 | Indels | 0 | Gaps | 0 |
|---------|----|--------------|---|------------|---|--------|---|------|---|
|---------|----|--------------|---|------------|---|--------|---|------|---|

Db 190 LNSKIAFKIVSOEPA 204
 QY 1 LNSKIAFKIVSOEPA 15

RESULT 2
 ENTRY IJBOG1 #type complete
 TITLE desmoglein 1 precursor - bovine
 ALTERNATE_NAMES desmoglein B0GM
 ORGANISM #formal_name Bos primigenius taurus #common_name cattle
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS
 REFERENCE S14603: A38872: A37785: S38721: A48173: S24412
 S14603
 #authors Koch, P.J.: Goldschmidt, M.D.: Zimbelmann, R.: Franke, W.W.
 #submission Submitted to the EMBL Data Library, March 1991
 #description Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.

#accession S14603
 #molecule_type mRNA
 #residues 1-1043 #label KOC
 REFERENCE #cross-references EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307
 A38872
 #authors Koch, P.J.: Goldschmidt, M.D.: Walsh, M.J.: Zimbelmann, R.: Franke, W.W.

#journal Eur. J. Cell Biol. (1991) 55:200-208
 #title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.

#cross-references M0ID:92037656
 #accession A38872
 #molecule_type mRNA
 #residues 1-87:968-1043 #label KO2
 REFERENCE #cross-references GB:S64268; GB:S64270
 A37785

#authors Goodwin, L.: Hill, J.E.: Raynor, K.: Raszi, L.: Manabe, M.: Cowin, P.

#journal Biochem. Biophys. Res. Commun. (1990) 173:1224-1230
 #title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.

#cross-references M0ID:91097553
 #accession A37785
 #molecule_type mRNA
 #residues 44-123,'V',125-493 #label GOO
 REFERENCE #cross-references GB:M58165; NID:g162966; PIDN:AAA62709.1; PID:g552318
 S38721

#authors Zimbelmann, R.
 #submission Submitted to the EMBL Data Library, February 1991
 #accession S38721
 #molecule_type mRNA
 #residues 44-1043 #label ZIM
 #cross-references EMBL:X57784; NID:g436061; PIDN:CAA40930.1; PID:g436062

REFERENCE A48173
 #authors Koch, P.J.: Walsh, M.J.: Schmelz, M.: Goldschmidt, M.D.: Zimbelmann, R.: Franke, W.W.
 #journal Eur. J. Cell Biol. (1990) 53:1-12
 #title Identification of desmoglein, a constitutive desmosomal glycoprotein, as a member of the cadherin family of cell adhesion molecules.

#cross-references M0ID:91168965
 #accession A48173
 #molecule_type mRNA
 #residues 44-1001,'AQP5AT' #label KO3
 REFERENCE #cross-references GB:X57784
 S38721

#note this sequence has been revised in references A38872 and S38721

GENETICS
 #gene DSG1
 CLASSIFICATION #superfamily cadherin: cadherin repeat homology
 #keywords calcium binding; cell adhesion; duplication; glycoprotein;

FEATURE
 1-23
 24-49
 50-1043
 50-548
 52-157
 150-269
 272-385
 392-491
 549-574
 575-1043
 846-875
 906-905
 906-933
 934-962
 963-1012
 110

transmembrane protein

SUMMARY
 #length 1043 #molecular_weight 112242 #checksum 6897

Query Match
 Best Local Similarity 80.0%; Pred. No. 2,27e-07;
 Matches 12: Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 LNSKIAFKIVSOEPA 204
 QY 1 LNSKIAFKIVSOEPA 15

RESULT 3
 ENTRY ITHUG1 #type complete
 TITLE desmoglein 1 precursor - human
 ALTERNATE_NAMES desmosomal glycoprotein I
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS
 REFERENCE S16906: A39706: A61254: S16158
 S16906
 #authors Buxton, R.S.
 #submission Submitted to the EMBL Data Library, November 1990
 #accession S16906
 #molecule_type mRNA
 #residues 1-1049 #label BUX
 REFERENCE #cross-references EMBL:X56654; NID:g30505; PIDN:CAA39976.1; PID:g30506
 A39706

#authors Wheeler, G.N.: Parker, A.E.: Thomas, C.L.: Attalio, P.: Poynter, D.: Arnemann, J.: Rutman, A.J.: Pidsley, S.C.: Watt, F.M.: Rees, D.A.: Buxton, R.S.: Magee, A.I.

#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4786-4800
 #title Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.

#cross-references M0ID:91271279
 #accession A39706
 #molecule_type mRNA
 #residues 24-1049 #label WHE
 REFERENCE #cross-references GB:X56654
 A61254

#authors Nilles, L.A.: Parry, D.A.D.: Powers, E.E.: Angst, B.D.: Wagner, R.M.: Green, K.J.
 #journal J. Cell Sci. (1991) 99:809-821
 #title Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.

#cross-references M0ID:92121251
 #accession A61254
 #molecule_type mRNA
 #residues 26-1049 #label NIL

REFERENCE A61279
 #authors Wheeler, G.N.: Buxton, R.S.: Parker, A.E.: Arnemann, J.: Rees, D.A.: King, I.A.: Magee, A.I.

#journal Biochem. Soc. Trans. (1991) 19:1060-1064
#title Desmosomal glycoproteins I, II and III: novel members of the
cadherin superfamily.
#cross-references MUID:92175187
#accession A61279
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-55 ##label WH3

GENETICS
#gene GDB:DSG1
#cross-references GDB:126563; OMIM:125670
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein

FEATURE
1-23 #domain signal sequence #status predicted #label SIG\
24-49 #domain propeptide #status predicted #label PRO\
50-1049 #product desmoglein #status predicted #label MAR\
50-548 #domain extracellular #status predicted #label EXT\
52-157 #domain cadherin repeat homology #label CR1\
160-269 #domain cadherin repeat homology #label CR2\
272-385 #domain cadherin repeat homology #label CR3\
392-493 #domain cadherin repeat homology #label CR4\
509-530 #region serine/threonine-rich\
549-569 #domain transmembrane #status predicted #label TMN\
572-1049 #domain intracellular #status predicted #label INT\
840-869 #domain desmoglein repeat #label DG1\
870-899 #domain desmoglein repeat #label DG2\
900-927 #domain desmoglein repeat #label DG3\
928-956 #domain desmoglein repeat #label DG4\
969-1019 #region glycine/serine-rich\
110,180 #binding_site carbohydrate (Asn) (covalent) #status
predicted

SUMMARY
#length 1049 #molecular-weight 113715 #checksum 4482

Query Match 90.6%; Score 87; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 2,27e-07;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 LNSKIAFKIIVSOEPA 204
1 LNSKIAFKIIVSOEPA 15

OY 1 LNSKIAFKIIVSOEPA 15

RESULT 4
ENTRY S38673 #type complete
TITLE desmoglein 2 - human
ALTERNATE_NAMES #formal_name Homo sapiens #common_name man
ORGANISM 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
DATE 20-Aug-1999
ACCESSIONS S38673; B38872
REFERENCE S38673
#authors Zimbelmann, R.
#submission submitted to the EMBL Data Library, September 1993
#accession S38673
#status preliminary
#molecule_type mRNA
#residues 1-1117 ##label ZIM
#cross-references EMBL:Z26317; NID:9416177; PIDN:CAA81226.1;
PID:9416178

REFERENCE A38872
#authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.;
Franko, W.W.
#journal Eur. J. Cell Biol. (1991) 55:200-208
#title Complete amino acid sequence of the epidermal desmoglein
precursor polypeptide and identification of a second type
of desmoglein gene.
#cross-references MUID:92037656
#accession B38872
#molecule_type mRNA
#residues 777-1117 ##label KOC

##cross-references GB:S64273

GENETICS
#gene GDB:DSG2
#cross-references GDB:128808; OMIM:125671
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
membrane protein

FEATURE
51-158 #domain cadherin repeat homology #label CR1\
161-271 #domain cadherin repeat homology #label CR2\
SUMMARY #length 1117 #molecular-weight 122384 #checksum 7660

Query Match 80.2%; Score 77; DB 2; Length 1117;
Best Local Similarity 73.3%; Pred. No. 6,37e-05;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 191 LNSKISYRIVSLEPA 205
1 LNSKIAFKIIVSOEPA 15

OY 1 LNSKIAFKIIVSOEPA 15

RESULT 5
ENTRY T02501 #type complete
TITLE hypothetical protein T19C21.7 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
14-May-1999
ACCESSIONS T02501
REFERENCE 214676
#authors Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
J.C.
#submission submitted to the EMBL Data Library, August 1998
#description Arabidopsis thaliana chromosome II BAC T19C21 genomic
sequence.
#accession T02501
#status translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-1421 ##label ROU
#cross-references EMBL:AC004683; NID:93395421; PID:93395428
##experimental_source cultivar Columbia

GENETICS
#map_position 2
#introns 52/2; 107/1; 148/2; 191/3; 212/2; 1317/3; 1346/3
#note T19C21.7
SUMMARY #length 1421 #molecular-weight 154325 #checksum 9888

Query Match 63.5%; Score 61; DB 2; Length 1421;
Best Local Similarity 40.0%; Pred. No. 2,63e-01;
Matches 6; Conservative 8; Mismatches 1; Indels 0; Gaps 0;

Db 262 VDSKISYEITONPG 276
1 LNSKIAFKIIVSOEPA 15

OY 1 LNSKIAFKIIVSOEPA 15

RESULT 6
ENTRY IUMSC2 #type complete
TITLE E-cadherin precursor, epithelial - mouse
ALTERNATE_NAMES uvomorulin
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
22-Jun-1999
ACCESSIONS S04528; S03160; I49565; S48735
REFERENCE S04528
#authors Nagafuchi, A.; Shirayoshi, Y.; Okazaki, K.; Yasuda, K.;
Takeichi, M.
#journal Nature (1987) 329:341-343
#title Transformation of cell adhesion properties by exogenously
introduced E-cadherin cDNA.

#cross-references MUID:87315445
#accession S04528
#molecule-type mRNA
#residues 1-412,'V',414-884 #label NAG
#cross-references EMBL:X06115
REFERENCE S03160
#authors Ringwald, M.; Schuh, R.; Vestweber, D.; Elstetter, H.; Lottspeich, F.; Engel, J.; Doebe, R.; Jaehnig, F.; Epplen, J.; Mayer, S.; Mueller, C.; Kemler, R.
#journal EMBO J. (1987) 6:3647-3653
#title The structure of cell adhesion molecule uvomorulin. Insights into the molecular mechanism of Ca(2+)-dependent cell adhesion.
#cross-references MUID:88111553
#accession S03160
#molecule-type mRNA
#residues 157-884 #label RIN
#cross-references EMBL:X06339
#note mature protein, was confirmed by protein sequencing
REFERENCE I49565
#authors Behrens, J.; Loswick, O.; Klein-Hitpass, L.; Birchmeier, W.; Proc. Natl. Acad. Sci. U.S.A. (1991) 88:11495-11499
#journal The E-cadherin promoter: Functional analysis of a G-C-rich region and an epithelial cell-specific palindromic regulatory element.
#cross-references MUID:92107977
#accession I49565
#status preliminary: translated from GB/EMBL/DBJ
#molecule-type DNA
#residues 1-15 #label RES
#cross-references GB:M01449; NID:g192325; PIDN:AAA37352.1; PID:g192326
REFERENCE S48735
#authors Tong, K.I.; Yau, P.; Overduin, M.; Bagby, S.; Portumb, T.; Takeichi, M.; Ikura, M.
#journal FEBS Lett. (1994) 352:318-322
#title Purification and spectroscopic characterization of a recombinant amino-terminal polypeptide fragment of mouse epithelial cadherin.
#cross-references MUID:95010732
#accession S48735
#status preliminary
#molecule-type protein
#residues 156-300 #label TON
COMMENT Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in the sorting of different cell types during morphogenesis.

GENETICS
#gene E-cadherin
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE
1-77 #domain signal sequence #status predicted #label SIG
28156 #domain propeptide #status predicted #label PRO
157-884 #product E-cadherin, epithelial #status experimental
#label MAT
157-699 #domain extracellular #status predicted #label EXT
159-264 #domain cadherin repeat homology #label CR1
234-239 #region cadherin binding #status predicted
267-377 #domain cadherin repeat homology #label CR2
380-488 #domain cadherin repeat homology #label CR3
489-597 #domain cadherin repeat homology #label CR4
598-702 #domain cadherin repeat homology #label CR5
703-733 #domain transmembrane #status predicted #label TM
734-884 #domain intracellular #status predicted #label INT
842-855 #domain serine-rich
560,659 #binding-site carbohydrate (Asn) (covalent) #status predicted
#length 884 #molecular-weight 98255 #checksum 2268

SUMMARY
Query Match 62.5%; Score 60; DB 1; Length 884;
Best Local Similarity 61.5%; Pred. No. 4.27e-01;

Matches 8: Conservative 3: Mismatches 2: Indels 0: Gaps 0:
Db 299 NAAIATVSDP 311
1: 11: 1111:1
Qy 2 NSKIAKIVSOEP 14

RESULT 7
ENTRY S34438 #type complete
TITLE uvomorulin - mouse
ORANISM #formal_name Mus musculus #common_name house mouse
DATE 22-Nov-1993 #sequence_revision 03-Nov-1995 #text_change 23-May-1997
ACCESSIONS S34438
REFERENCE S34438
#authors Ringwald, M.; Barbaulet, H.; Schmidt, C.; Kemler, R.
#journal Nucleic Acids Res. (1991) 19:6533-6539
#title The structure of the gene coding for the mouse cell adhesion molecule uvomorulin.
#cross-references MUID:92093614
#accession S34438
#status preliminary: nucleic acid sequence not shown:
#molecule-type DNA
#residues 1-884 #label RIN
#cross-references EMBL:X60975
#note the nucleotide sequence was submitted to the EMBL Data Library, August 1992

GENETICS
#introns 16/3: 57/1: 131/3: 179/3: 231/3: 280/1: 338/3: 381/3: 442/3:
524/2: 573/1: 648/1: 724/1: 767/3: 815/3
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
FEATURE 380-488
SUMMARY #domain cadherin repeat homology #label CR3
#length 884 #molecular-weight 98283 #checksum 2125

Query Match 62.5%; Score 60; DB 2; Length 884;
Best Local Similarity 61.5%; Pred. No. 4.27e-01;
Matches 8: Conservative 3: Mismatches 2: Indels 0: Gaps 0:
Db 299 NAAIATVSDP 311
1: 11: 1111:1
Qy 2 NSKIAKIVSOEP 14

RESULT 8
ENTRY ITHUCN #type complete
TITLE cadherin 2 precursor - human
ALTERNATE_NAMES N-cadherin; neuronal cadherin
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
ACCESSIONS A38870; S11487; J00751; S13799
REFERENCE A38870
#authors Reid, R.A.
#submsion submitted to the EMBL Data Library, November 1990
#accession A38870
#molecule-type mRNA
#residues 1-906 #label RE1
#cross-references EMBL:X54315; NID:g34998; PIDN:CA38213.1; PID:g34999
REFERENCE S11487
#authors Reid, R.A.; Hemperly, J.J.
#journal Nucleic Acids Res. (1990) 18:5896
#title Human N-cadherin: nucleotide and deduced amino acid sequence.
#cross-references MUID:91016546
#accession S11487
#molecule-type mRNA
#residues 1-340,'N',342-698,'R',700-704,'F',706-906 #label RE2
#cross-references EMBL:X54315
#note this sequence has been revised in reference A38870
REFERENCE J00751
#authors Walsh, F.S.; Barton, C.H.; Putt, W.; Moore, S.E.; Kelsell, D.; Spurr, N.; Goodfellow, P.N.

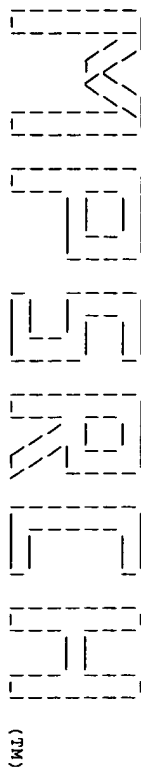
REFERENCE 564144
#authors Escribano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.
#submission Submitted to the Protein Sequence Database, May 1996
#accession S64146
#molecule_type DNA
#residues 1-1264 #label ESC
#cross-references EMBL:272655; NID:g1322701; PID:e243361; PID:g1322702;
MIPS:YGL133Y
#experimental_source strain S288C
REFERENCE
#authors Escribano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.
#journal Yeast (1996) 12:887-892
#title Sequence analysis of a 14.6 kb DNA fragment of *Saccharomyces cerevisiae* chromosome VII reveals SEC27, SSM1b, a putative S-adenosylmethionine-dependent enzyme and six new open reading frames.
#cross-references MUID:96437978
#accession S71739
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-1264 #label ESM
#cross-references EMBL:X82670
#note the nucleotide sequence was submitted to the EMBL Data Library, October 1995
GENETICS
#map_position 7L
KEYWORDS transmembrane protein
FEATURE
549-565 #domain transmembrane #status predicted #label TM
SUMMARY #length 1264 #molecular-weight 145642 #checksum 9929
Query Match 59.4%; Score 57; DB 2; Length 1264;
Best Local Similarity 42.9%; Pred. No. 1.77e+00;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
DB 524 SDRVEKIVHDDPA 537
QY 2 NSKIAFKIVSOEPA 15
RESULT 12
ENTRY IJFTM #type complete
TITLE cadherin-related tumor suppressor precursor - fruit fly
ORGANISM *Drosophila melanogaster*
#formal_name *Drosophila melanogaster*
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Feb-1997
ACCESSION A41087
REFERENCE
#authors Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Blessmann, R.; Bryant, P.J.; Goodman, C.S.
#journal Cell (1991) 67:853-868
#title The fat tumor suppressor gene in *Drosophila* encodes a novel member of the cadherin gene superfamily.
#cross-references MUID:92069752
#accession A41087
#molecule_type mRNA
#residues 143-485;1279-5147 #label MAH
#cross-references GB:M80537
#accession B41087
#molecule_type DNA
#residues 1-142;487-1278 #label MAZ
#cross-references GB:M80537
#note 1229-Gly and 1233-Ser were also found
GENETICS
#gene fat
#cross-references FlyBase:FBgn0001075
CLASSIFICATION #superfamily cadherin-related tumor suppressor; cadherin
KEYWORDS repeat homology; EGF homology
calcium binding; cell adhesion; duplication; transmembrane protein
FEATURE
1-35 #domain signal sequence #status predicted #label SIG\

36-5147
36-4583
51-156
159-270
271-382
390-494
407-509
602-708
718-822
948-1049
952-1133
1156-1278
1281-1384
1387-1489
1492-1601
1607-1713
1717-1823
1826-1922
1925-2027
2028-2167
2169-2278
2281-2384
2387-2491
2494-2596
2599-2703
2707-2810
2813-2913
2915-3013
3014-3124
3127-3229
3232-3334
3337-3439
3442-3545
3548-3651
3654-3756
3954-4010
4017-4048
4056-4089
4096-4127
4484-4609
4610-5147
#product cadherin-related tumor suppressor #status predicted #label MAT
#domain extracellular #status predicted #label EXT
#domain cadherin repeat homology #label CR1
#domain cadherin repeat homology #label CR2
#domain cadherin repeat homology #label CR3
#domain cadherin repeat homology #label CR4
#domain cadherin repeat homology #label CR5
#domain cadherin repeat homology #label CR6
#domain cadherin repeat homology #label CR7
#domain cadherin repeat homology #label CR8
#domain cadherin repeat homology #label CR9
#domain cadherin repeat homology #label CR10
#domain cadherin repeat homology #label CR11
#domain cadherin repeat homology #label CR12
#domain cadherin repeat homology #label CR13
#domain cadherin repeat homology #label CR14
#domain cadherin repeat homology #label CR15
#domain cadherin repeat homology #label CR16
#domain cadherin repeat homology #label CR17
#domain cadherin repeat homology #label CR18
#domain cadherin repeat homology #label CR19
#domain cadherin repeat homology #label CR20
#domain cadherin repeat homology #label CR21
#domain cadherin repeat homology #label CR22
#domain cadherin repeat homology #label CR23
#domain cadherin repeat homology #label CR24
#domain cadherin repeat homology #label CR25
#domain cadherin repeat homology #label CR26
#domain cadherin repeat homology #label CR27
#domain cadherin repeat homology #label CR28
#domain cadherin repeat homology #label CR29
#domain cadherin repeat homology #label CR30
#domain cadherin repeat homology #label CR31
#domain cadherin repeat homology #label CR32
#domain cadherin repeat homology #label CR33
#domain EGF homology #label EG1
#domain EGF homology #label EG2
#domain EGF homology #label EG3
#domain EGF homology #label EG4
#domain transmembrane #status predicted #label TM
#domain intracellular #status predicted #label INT
#length 5147 #molecular-weight 564895 #checksum 6994
Query Match 59.4%; Score 57; DB 1; Length 5147;
Best Local Similarity 30.0%; Pred. No. 1.77e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
DB 1520 LNKVSYAISKOE 1533
QY 1 LNSKIAFKIVSOE 14
RESULT 13
ENTRY F70126 #type complete
TITLE translocation elongation factor P (efp) homolog - Lyme disease spirochete
ORGANISM *Borrelia burgdorferi* #common_name Lyme disease spirochete
DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 29-Sep-1999
ACCESSION F70126
REFERENCE F70126
#authors Fraser, C.M.; Castens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwin, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Melman, J.; Uterback, T.; Watthey, L.; McDonald, L.; Artach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.

#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
#cross-references MUID:98065943
#accession F70126
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-192 #label KLE
##cross-references GB:AE001132; GB:AE000763; NID:g2688107;
PIDN:AAC66610.1; PID:g2688116; TIGR:BB0214
##experimental_source strain B31
CLASSIFICATION #superfamily translation elongation factor EF-P
SUMMARY #length 192 #molecular-weight 21413 #checksum 4881
Query Match 57.3%; Score 55; DB 2; Length 192;
Best Local Similarity 53.3%; Pred. No. 4,45e+00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Db 130 LAPKAFVEVEEA 144
OY 1 LNSKIAFKIVSOEPA 15
RESULT 14
ENTRY T05921 #type fragment
TITLE N-carbamyl-L-amino acid amidohydrolase homolog - barley
(fragment)
ORGANISM #formal_name Hordeum vulgare #common_name barley
DATE 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change
T05921
215411
ACCESSIONS Hess, W.R.; Golz, R.R.; Boerner, T.
REFERENCE #journal Plant Sci. (1998) 133:191-201
#title Analysis of randomly selected cDNAs reveals the expression
of stress- and defence related genes in the barley mutant
alostrians.
#accession T05921
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-108 #label HES
##cross-references EMBL:AJ22776; NID:e1203983; PID:e1203984
##experimental_source cv. Halsa, leaf
SUMMARY #length 108 #checksum 4694
Query Match 56.3%; Score 54; DB 2; Length 108;
Best Local Similarity 45.5%; Pred. No. 7,00e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 15 LEFKINDPP 25
OY 5 IAFKIVSOEPA 15
RESULT 15
ENTRY IJB0CN #type fragment
TITLE N-cadherin precursor - bovine (fragment)
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
22-Jun-1999
ACCESSIONS S11693
REFERENCE S11693
#authors Liaw, C.W.; Cannon, C.; Power, M.D.; Kibonoka, P.K.; Rubin,
L.L.
#journal EMBO J. (1990) 9:2701-2708
#title Identification and cloning of two species of cadherins in
bovine endocelial cells.
#cross-references MUID:90360979
#accession S11693
##molecule_type mRNA
##residues 1-877 #label LIA
##cross-references EMBL:X53615; NID:g164; PID:CAA37677.1; PID:g664894

COMMENT Cadherins mediate calcium-dependent intercellular adhesion, and are
thought to be involved in the sorting of different cell types
during morphogenesis.
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein
FEATURE
1-130
#domain propeptide (fragment) #status predicted #label
PRO
131-877
131-885 #product N-cadherin #status predicted #label MAT
133-238 #domain extracellular #status predicted #label EXT
208-213 #domain cadherin repeat homology #label CR1
241-353 #region cadherin binding #status predicted
356-468 #domain cadherin repeat homology #label CR2
471-576 #domain cadherin repeat homology #label CR3
577-685 #domain cadherin repeat homology #label CR4
686-717 #domain cadherin repeat homology #label CR5
718-877 #domain transmembrane #status predicted #label TM1
836-849 #domain intracellular #status predicted #label INT
161,244,296,373, #binding_site carbohydrate (Asn) (covalent) #status
543,593,622,663 predicted
SUMMARY #length 877 #checksum 8685
Query Match 56.3%; Score 54; DB 1; Length 877;
Best Local Similarity 40.0%; Pred. No. 7,00e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
Db 271 LNCMLRYRIISQAPS 285
OY 1 LNSKIAFKIVSOEPA 15

Search completed: Sat May 13 07:10:30 2000
Job time : 9 secs.


 (TM)

Release 3.1A John F. Collins, Bioinformatics Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSEARCH protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 07:05:41 2000; Maspar time: 79.42 seconds

Tabular output not generated. 5.752 Million cell updates/sec

Title: >US-08-991-628-3

Description: (1-15) from US08991628.pep

Perfect Score: 96

Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: PAM 150

Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot38

1:swissprot

Statistics: Mean 26.327; Variance 28.416; scale 0.926

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|-------------|------------------------|-----------|
| 1 | 96 | 100.0 | 999 | 1 | DSG3_HUMAN | DESMOGLEIN 3 PRECURSOR | 1.17e-11 |
| 2 | 87 | 90.6 | 1043 | 1 | DSG1_BOVIN | DESMOGLEIN 1 PRECURSOR | 5.08e-09 |
| 3 | 87 | 90.6 | 1049 | 1 | DSG1_HUMAN | DESMOGLEIN 1 PRECURSOR | 5.08e-09 |
| 4 | 77 | 80.2 | 1117 | 1 | DSG3_HUMAN | DESMOGLEIN 2 PRECURSOR | 3.24e-06 |
| 5 | 60 | 62.5 | 884 | 1 | CAD1_MOUSE | EPITHELIAL-CADHERIN PR | 7.53e-02 |
| 6 | 60 | 62.5 | 906 | 1 | CAD2_HUMAN | NEURAL-CADHERIN PRECUR | 3.79e-01 |
| 7 | 59 | 61.5 | 385 | 1 | YOX1_YEAST | HOMEOBOX PROTEIN YOX1 | 1.30e-01 |
| 8 | 57 | 59.4 | 1264 | 1 | YGN3_YEAST | HYPOTHETICAL 145.6 KD | 3.79e-01 |
| 9 | 57 | 59.4 | 5147 | 1 | FAT_DROME | CADHERIN-RELATED TUMOR | 3.79e-01 |
| 10 | 55 | 57.3 | 192 | 1 | EPF_BOBBU | ELONGATION FACTOR P (E | 1.08e+00 |
| 11 | 54 | 56.3 | 877 | 1 | CAD2_BOVIN | NEURAL-CADHERIN PRECUR | 1.79e+00 |
| 12 | 54 | 56.3 | 882 | 1 | CAD1_HUMAN | EPITHELIAL-CADHERIN PR | 1.79e+00 |
| 13 | 54 | 56.3 | 905 | 1 | CAD2_XENLA | BLASTOMERE-CADHERIN PR | 1.79e+00 |
| 14 | 54 | 56.3 | 906 | 1 | CAD2_MOUSE | NEURAL-CADHERIN PRECUR | 1.79e+00 |
| 15 | 54 | 56.3 | 1146 | 1 | AS10_YEAST | ASK10 PROTEIN | 1.79e+00 |
| 16 | 54 | 56.3 | 2233 | 1 | RRPL_PI3H4 | RNA POLYMERASE BETA SU | 2.97e+00 |
| 17 | 53 | 55.2 | 887 | 1 | CAD1_CHICK | EPITHELIAL-CADHERIN PR | 2.97e+00 |
| 18 | 53 | 55.2 | 2048 | 1 | RRPL_SENDE | RNA POLYMERASE BETA SU | 2.97e+00 |
| 19 | 53 | 55.2 | 2228 | 1 | RRPL_SENDE | RNA POLYMERASE BETA SU | 2.97e+00 |
| 20 | 53 | 55.2 | 2228 | 1 | RRPL_SENDE | RNA POLYMERASE BETA SU | 2.97e+00 |
| 21 | 53 | 55.2 | 2228 | 1 | RRPL_SENDE | RNA POLYMERASE BETA SU | 2.97e+00 |
| 22 | 52 | 54.2 | 109 | 1 | PRV1_SALSA | PARVALBUMIN BETA 1 (CL | 4.88e+00 |
| 23 | 52 | 54.2 | 188 | 1 | RUVA_THEMEA | HOLLIDAY JUNCTION DNA | 4.88e+00 |

| | | | | | | | |
|----|----|------|------|---|------------|-------------------------|----------|
| 24 | 52 | 54.2 | 732 | 1 | CAD1_CHICK | B-CADHERIN PRECURSOR (| 4.88e+00 |
| 25 | 52 | 54.2 | 821 | 1 | CAN3_RAT | CALPAIN P94, LARGE (CA | 4.88e+00 |
| 26 | 52 | 54.2 | 821 | 1 | CAN3_MOUSE | SURFACE PRESENTATION O | 4.88e+00 |
| 27 | 51 | 53.1 | 293 | 1 | SPAO_SHFL | HYPOTHETICAL 74.5 KD P | 7.95e+00 |
| 28 | 51 | 53.1 | 644 | 1 | YEAG_ECOLI | ISOCITRATE DEHYDROGENA | 7.95e+00 |
| 29 | 51 | 53.1 | 741 | 1 | IDH2_VIBRI | HYPOTHETICAL PROTEIN M | 7.95e+00 |
| 30 | 51 | 53.1 | 1783 | 1 | Y468_MYCE | ASPARTIC PROTEINASE IN | 1.29e+01 |
| 31 | 50 | 52.1 | 217 | 1 | IAP2_SOLTU | CARBONYL REDUCTASE (NA | 1.29e+01 |
| 32 | 50 | 52.1 | 276 | 1 | DHCA_RAT | PROBABLE METHYLTRANSFER | 1.29e+01 |
| 33 | 50 | 52.1 | 312 | 1 | PRMA_STMAU | VACB SHFL | 1.29e+01 |
| 34 | 50 | 52.1 | 760 | 1 | VACB_SHFL | TER1_CAEEL | 1.29e+01 |
| 35 | 50 | 52.1 | 809 | 1 | TER1_CAEEL | GENOME POLYPROTEIN 1 (| 1.29e+01 |
| 36 | 50 | 52.1 | 2410 | 1 | POL1_BAYM | GENOME POLYPROTEIN 1 (| 1.29e+01 |
| 37 | 50 | 52.1 | 2412 | 1 | POL1_BAYM | GENOME POLYPROTEIN 1 (| 1.29e+01 |
| 38 | 49 | 51.0 | 266 | 1 | ERM1_BACFR | RNA ADENINE N-6-METHY | 2.06e+01 |
| 39 | 49 | 51.0 | 306 | 1 | LUXD_PHOPO | ACYL TRANSFERASE (EC 2 | 2.06e+01 |
| 40 | 49 | 51.0 | 318 | 1 | CYF_OENHO | APOCYTOCHROME F PRECUR | 2.06e+01 |
| 41 | 49 | 51.0 | 320 | 1 | CYF_TOBAC | APOCYTOCHROME F PRECUR | 2.06e+01 |
| 42 | 49 | 51.0 | 320 | 1 | CYF_VICPA | APOCYTOCHROME F PRECUR | 2.06e+01 |
| 43 | 49 | 51.0 | 320 | 1 | CYF_PEA | APOCYTOCHROME F PRECUR | 2.06e+01 |
| 44 | 49 | 51.0 | 810 | 1 | TER2_CAEEL | TRANSITIONAL ENDOPLASM | 2.06e+01 |
| 45 | 49 | 51.0 | 916 | 1 | CAD4_HUMAN | RETINAL-CADHERIN PRECU | 2.06e+01 |

ALIGNMENTS

RESULT 1
 ID DSG3_HUMAN STANDARD: PRT; 999 AA.

AC P32926;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).

GN DSG3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Homidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 92069753.

RA Amagai M., Klaus-Kovtun V., Stanley J.R.;

RT "Autoantibodies against a novel epithelial cadherin in pemphigus

KT vulgaris, a disease of cell adhesion.";

RL Cell 67:869-877(1991).

CC -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.

CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE

CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND

CC CARCINOMAS.

CC -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS

CC (POTENTIAL).

CC -!- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN

CC DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE

CC LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES

CC AGAINST DSG3.

CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>

CC or send an email to license@sdb.ch).

CC EMBL: M76482; AAA60230.1; -

CC PIR: A41088; IJH0G3.

CC HSSP: P09803; 1EDH.

CC MIM: 169615; -

CC PROSITE: PS00232; CADHERIN: 3.

CC PFM: PF00028; cadherin: 4.

CC Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;

KW Calcium-binding: Repeat.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 49 POTENTIAL.
 FT CHAIN 30 999 DESMOGLEIN 3.
 FT DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 616 640 POTENTIAL.
 FT DOMAIN 641 999 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 50 158 CADHERIN 1.
 FT REPEAT 159 268 CADHERIN 2.
 FT REPEAT 269 383 CADHERIN 3.
 FT REPEAT 386 499 CADHERIN 4.
 FT REPEAT 910 935 DESMOGLEIN REPEAT 1.
 FT REPEAT 936 966 DESMOGLEIN REPEAT 2.
 FT CARBOHYD 110 110 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 459 459 POTENTIAL.
 FT CARBOHYD 545 545 POTENTIAL.
 SQ SEQUENCE 999 AA: 107503 MW: 60479DD46AC219A1 CRC64:

Query Match 100.0% Score 96: DB 1: Length 999;
 Best Local Similarity 100.0% Pred. No. 1.17e-11;
 Matches 15: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Db 190 LNSKIAFKIVSOEPA 204
 Oy 1 LNSKIAFKIVSOEPA 15

RESULT 2
 ID DSG1_BOVIN STANDARD: PRT: 1043 AA.
 AC 003763:
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DGI).
 GN DSG1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-MUZZLE EPITHELIUM;
 RA Koch P.J., Goldschmidt M.D., Zimbelmann R., Franke W.W.;
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 44-1043 FROM N.A.
 RC TISSUE-MUZZLE EPITHELIUM;
 RA MEDLINE: 91168965.
 RA Koch P.J., Walsh M.J., Schmelz M., Goldschmidt M.D.,
 RA Zimbelmann R., Franke W.W.;
 RT "Identification of desmoglein, a constitutive desmosomal
 RT glycoprotein, as a member of the cadherin family of cell adhesion
 RT molecules.";
 RL Eur. J. Cell Biol. 53:1-12(1990).
 RN [3]
 RP REVISIONS, AND SEQUENCE OF 101-123.
 RX MEDLINE: 92037656.
 RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R.,
 RA Franke W.W.;
 RT "Complete amino acid sequence of the epidermal desmoglein precursor
 RT polypeptide and identification of a second type of desmoglein gene.";
 RL Eur. J. Cell Biol. 55:200-208(1991).
 RN [4]
 RP SEQUENCE OF 44-493 FROM N.A.
 RC MEDLINE: 91097553.
 RA Goodwin L., Hill J.E., Raynor K., Raszi L., Manabe M., Cowin P.;
 RT "Desmoglein shows extensive homology to the cadherin family of cell
 RT adhesion molecules.";
 RL Biochem. Biophys. Res. Commun. 173:1224-1230(1990).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOsome SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X58466; CAA41380.1; -;
 DR EMBL: X57784; CAA40930.1; -;
 DR EMBL: M58165; AAA62709.1; -;
 DR PIR: S14603; IYBOG1.
 DR HSSP: P09803; 1EDH.
 DR PROSITE: PS00232; CADHERIN; 2.
 DR PRAM: PF00028; cadherin; 3.
 DR KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
 KW Calcium-binding; Repeat.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 49 DESMOGLEIN 1.
 FT CHAIN 50 1043 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 50 548 POTENTIAL.
 FT TRANSMEM 549 573 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 574 1043 CADHERIN 1.
 FT REPEAT 50 158 CADHERIN 2.
 FT REPEAT 159 270 CADHERIN 3.
 FT REPEAT 271 385 CADHERIN 4.
 FT REPEAT 386 498 DESMOGLEIN REPEAT 1.
 FT REPEAT 419 845 DESMOGLEIN REPEAT 2.
 FT REPEAT 846 875 DESMOGLEIN REPEAT 3.
 FT REPEAT 876 905 DESMOGLEIN REPEAT 4.
 FT REPEAT 906 933 DESMOGLEIN REPEAT 5.
 FT REPEAT 934 962 GLY/SER-RICH.
 FT DOMAIN 963 1012 POTENTIAL.
 FT CARBOHYD 110 110 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 496 496 POTENTIAL.
 FT CONFLICT 124 124 I -> V (IN REF. 4).
 SQ SEQUENCE 1043 AA: 112243 MW: ADE46133P8B77C11 CRC64:

Query Match 90.6% Score 87: DB 1: Length 1043;
 Best Local Similarity 80.0% Pred. No. 5.08e-09;
 Matches 12: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

Db 190 LNSKIAFKIVSOEPA 204
 Oy 1 LNSKIAFKIVSOEPA 15

RESULT 3
 ID DSG1_HUMAN STANDARD: PRT: 1049 AA.
 AC 002413:
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DGI).
 GN DSG1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KERATINOCYTES;
 RX MEDLINE: 91271279.
 RA Wheeler G.N., Parker A.E., Thomas C.L., Ataliotis P., Poynter D.,
 RA Arneemann J., Rutman A.J., Pidsley S.C., Watt F.M., Rees D.A.,
 RA Buxton R.S., Magee A.I.;
 RT "Desmosomal glycoprotein DGI, a component of intercellular desmosome

```

RT junctions, is related to the cadherin family of cell adhesion
RT molecules."
RT Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL AND ESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56654; CAA39976.1; -.
DR PIR: S16906; IJHUG1.
DR HSSP: P09803; IEDH.
DR MIM: 125670; -.
DR PROSITE: PS00232; CADHERIN. 2.
DR PRAM: PF00028; cadherin. 4.
KM Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
FT SIGNAL 23
FT PROPEP 24
FT CHAIN 49
FT DOMAIN 50 1049
FT TRANSSEM 546 570
FT DOMAIN 571 1049
FT REPEAT 50 158
FT REPEAT 159 270
FT REPEAT 271 385
FT REPEAT 386 497
FT REPEAT 497 839
FT REPEAT 839 869
FT REPEAT 869 899
FT REPEAT 900 927
FT REPEAT 928 956
FT DOMAIN 969 1019
FT CARBOHYD 36 36
FT CARBOHYD 110 110
FT CARBOHYD 180 180
SQ SEQUENCE 1049 AA; 113715 MW; EEI125655B9D6619 CRC64;

Query Match 90.6%; Score 87; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 5.08e-09;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 LNSKIARIROEPS 204
OY 1 LNSKIARIKIVSOEPA 15

RESULT 4
ID DSG2_HUMAN STANDARD; PRT; 1117 AA.
AC 014126;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 2 PRECURSOR (HDG).
DSG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE-COLON CARCINOMA;
RX MEDLINE: 94192736.

```

```

RA Schaefer S., Koch P.J., Franke W.W.;
RT "Identification of the ubiquitous human desmoglein, Dsg2, and the
RT expression catalogue of the desmoglein subfamily of desmosomal
RT cadherins."
RL Exp. Cell Res. 211:391-399(1994).
RN 12
RP SEQUENCE OF 777-1117 FROM N.A.
RX MEDLINE: 92037656.
RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R., Franke W.W.;
RT "Complete amino acid sequence of the epidermal desmoglein precursor
RT polypeptide and identification of a second type of desmoglein gene."
RL Eur. J. Cell Biol. 55:200-208(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z26317; CAA81226.1; -.
DR HSSP: P15116; INCI.
DR MIM: 125671; -.
DR PRINTS: PR00205; CADHERIN.
DR PROSITE: PS00232; CADHERIN. 3.
DR PRAM: PF00028; cadherin. 4.
KM Cell adhesion; Signal; Transmembrane; Repeat; Signal;
KW Cytoskeleton; Calcium-binding.
FT SIGNAL 23
FT PROPEP 24
FT CHAIN 48
FT DOMAIN 49 1117
FT TRANSSEM 609 633
FT DOMAIN 634 1117
FT REPEAT 49 159
FT REPEAT 160 272
FT REPEAT 273 387
FT REPEAT 388 502
FT REPEAT 502 911
FT REPEAT 912 941
FT REPEAT 942 967
FT REPEAT 968 991
FT REPEAT 992 1020
FT REPEAT 1021 1050
FT CARBOHYD 111 111
FT CARBOHYD 181 181
FT CARBOHYD 308 308
FT CARBOHYD 461 461
FT CARBOHYD 513 513
SQ SEQUENCE 1117 AA; 122385 MW; 223B897FED70B289 CRC64;

Query Match 80.2%; Score 77; DB 1; Length 1117;
Best Local Similarity 73.3%; Pred. No. 3.24e-06;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 191 LNSKISRIYSLSPA 205
OY 1 LNSKIARIKIVSOEPA 15

RESULT 5
ID CADL_MOUSE STANDARD; PRT; 884 AA.
AC P09803; 061377;
DT 01-MAR-1989 (Rel. 10, Created)

```


01-MAR-1989 (Rel. 10, last sequence update)
 DT 15-JUL-1998 (Rel. 36, last annotation update)
 DE EPIHELIAL-CADHERIN PRECURSOR (E-CADHERIN) (UVOXOLIN) (ARC-1).
 GN CDH1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-1CR.
 RX MEDLINE: 87315445.
 RA Nagafuchi A., Shirayoshi Y., Okazaki K., Yasuda K., Takeichi M.;
 RT "Transformation of cell adhesion properties by exogenously introduced
 RT E-cadherin cDNA.";
 RL Nature 329:341-343(1987).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV.
 RX MEDLINE: 92093614.
 RA Ringwald M., Barbault H., Schmidt C., Kemler R.;
 RT "The structure of the gene coding for the mouse cell adhesion
 RT molecule uvomorulin.";
 RL Nucleic Acids Res. 19:6533-6539(1991).
 RN
 RP SEQUENCE OF 174-884 FROM N.A., AND SEQUENCE OF 157-181.
 RA Ringwald M., Schuh R., Vestweber D., Eistetter H., Lottspeich F.,
 RA Engel J., Diehl R., Jaehnis F., Epplein J., Mayer S., Mueller C.,
 RA Kemler R.;
 RT "The structure of cell adhesion molecule uvomorulin. Insights into
 RT the molecular mechanism of Ca²⁺-dependent cell adhesion.";
 RL EMBO J. 6:3647-3653(1987).
 RN
 RP SEQUENCE OF 1-15 FROM N.A.
 RX MEDLINE: 92107977.
 RA Behrens J., Loewerick O., Klein-Hitpass L., Birchmeier W.;
 RT "The E-cadherin promoter: functional analysis of a G.C-rich region
 RT and an epithelial cell-specific palindromic regulatory element.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:11495-11499(1991).
 RN
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 157-370.
 RX MEDLINE: 96176249.
 RA Negar B., Overduin M., Ikura M., Rini J.M.;
 RT "Structural basis of calcium-induced E-cadherin rigidification and
 RT dimerization.";
 RL Nature 380:360-364(1996).
 RN
 RP STRUCTURE BY NMR OF 157-260.
 RX MEDLINE: 96271285.
 RA Overduin M., Tong K.I., Kay C.M., Ikura M.;
 RT "1H, 15N and 13C resonance assignments and monomeric structure of the
 RT amino-terminal extracellular domain of epithelial cadherin.";
 RL J. Biomol. NMR 7:173-189(1996).
 CC
 CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES.
 CC
 CC -!- SUBUNIT: HOMODIMER.
 CC
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC
 CC -!- TISSUE SPECIFICITY: NON-NEURAL EPITHELIAL TISSUES.
 CC
 CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
 CC or send an email to license@isb-sdb.ch).
 CC
 CC EMBL: X06115; CAA29488.1; -;
 CC EMBL: X60961; CAA43292.1; -;
 CC EMBL: X60962; CAA43292.1; JOINED.

DR EMBL: X60963; CAA43292.1; JOINED.
 DR EMBL: X60964; CAA43292.1; JOINED.
 DR EMBL: X60965; CAA43292.1; JOINED.
 DR EMBL: X60966; CAA43292.1; JOINED.
 DR EMBL: X60967; CAA43292.1; JOINED.
 DR EMBL: X60968; CAA43292.1; JOINED.
 DR EMBL: X60969; CAA43292.1; JOINED.
 DR EMBL: X60970; CAA43292.1; JOINED.
 DR EMBL: X60971; CAA43292.1; JOINED.
 DR EMBL: X60972; CAA43292.1; JOINED.
 DR EMBL: X60973; CAA43292.1; JOINED.
 DR EMBL: X60974; CAA43292.1; JOINED.
 DR EMBL: X60975; CAA43292.1; JOINED.
 DR EMBL: X60976; CAA43292.1; JOINED.
 DR EMBL: X60977; CAA43292.1; JOINED.
 DR EMBL: X60978; CAA43292.1; JOINED.
 DR EMBL: X60979; CAA43292.1; JOINED.
 DR EMBL: X60980; CAA43292.1; JOINED.
 DR EMBL: X60981; CAA43292.1; JOINED.
 DR EMBL: X60982; CAA43292.1; JOINED.
 DR EMBL: X60983; CAA43292.1; JOINED.
 DR EMBL: X60984; CAA43292.1; JOINED.
 DR EMBL: X60985; CAA43292.1; JOINED.
 DR EMBL: X60986; CAA43292.1; JOINED.
 DR EMBL: X60987; CAA43292.1; JOINED.
 DR EMBL: X60988; CAA43292.1; JOINED.
 DR EMBL: X60989; CAA43292.1; JOINED.
 DR EMBL: X60990; CAA43292.1; JOINED.
 DR EMBL: X60991; CAA43292.1; JOINED.
 DR EMBL: X60992; CAA43292.1; JOINED.
 DR EMBL: X60993; CAA43292.1; JOINED.
 DR EMBL: X60994; CAA43292.1; JOINED.
 DR EMBL: X60995; CAA43292.1; JOINED.
 DR EMBL: X60996; CAA43292.1; JOINED.
 DR EMBL: X60997; CAA43292.1; JOINED.
 DR EMBL: X60998; CAA43292.1; JOINED.
 DR EMBL: X60999; CAA43292.1; JOINED.
 DR EMBL: X61000; CAA43292.1; JOINED.
 DR EMBL: X61001; CAA43292.1; JOINED.
 DR EMBL: X61002; CAA43292.1; JOINED.
 DR EMBL: X61003; CAA43292.1; JOINED.
 DR EMBL: X61004; CAA43292.1; JOINED.
 DR EMBL: X61005; CAA43292.1; JOINED.
 DR EMBL: X61006; CAA43292.1; JOINED.
 DR EMBL: X61007; CAA43292.1; JOINED.
 DR EMBL: X61008; CAA43292.1; JOINED.
 DR EMBL: X61009; CAA43292.1; JOINED.
 DR EMBL: X61010; CAA43292.1; JOINED.
 DR EMBL: X61011; CAA43292.1; JOINED.
 DR EMBL: X61012; CAA43292.1; JOINED.
 DR EMBL: X61013; CAA43292.1; JOINED.
 DR EMBL: X61014; CAA43292.1; JOINED.
 DR EMBL: X61015; CAA43292.1; JOINED.
 DR EMBL: X61016; CAA43292.1; JOINED.
 DR EMBL: X61017; CAA43292.1; JOINED.
 DR EMBL: X61018; CAA43292.1; JOINED.
 DR EMBL: X61019; CAA43292.1; JOINED.
 DR EMBL: X61020; CAA43292.1; JOINED.
 DR EMBL: X61021; CAA43292.1; JOINED.
 DR EMBL: X61022; CAA43292.1; JOINED.
 DR EMBL: X61023; CAA43292.1; JOINED.
 DR EMBL: X61024; CAA43292.1; JOINED.
 DR EMBL: X61025; CAA43292.1; JOINED.
 DR EMBL: X61026; CAA43292.1; JOINED.
 DR EMBL: X61027; CAA43292.1; JOINED.
 DR EMBL: X61028; CAA43292.1; JOINED.
 DR EMBL: X61029; CAA43292.1; JOINED.
 DR EMBL: X61030; CAA43292.1; JOINED.
 DR EMBL: X61031; CAA43292.1; JOINED.
 DR EMBL: X61032; CAA43292.1; JOINED.
 DR EMBL: X61033; CAA43292.1; JOINED.
 DR EMBL: X61034; CAA43292.1; JOINED.
 DR EMBL: X61035; CAA43292.1; JOINED.
 DR EMBL: X61036; CAA43292.1; JOINED.
 DR EMBL: X61037; CAA43292.1; JOINED.
 DR EMBL: X61038; CAA43292.1; JOINED.
 DR EMBL: X61039; CAA43292.1; JOINED.
 DR EMBL: X61040; CAA43292.1; JOINED.
 DR EMBL: X61041; CAA43292.1; JOINED.
 DR EMBL: X61042; CAA43292.1; JOINED.
 DR EMBL: X61043; CAA43292.1; JOINED.
 DR EMBL: X61044; CAA43292.1; JOINED.
 DR EMBL: X61045; CAA43292.1; JOINED.
 DR EMBL: X61046; CAA43292.1; JOINED.
 DR EMBL: X61047; CAA43292.1; JOINED.
 DR EMBL: X61048; CAA43292.1; JOINED.
 DR EMBL: X61049; CAA43292.1; JOINED.
 DR EMBL: X61050; CAA43292.1; JOINED.
 DR EMBL: X61051; CAA43292.1; JOINED.
 DR EMBL: X61052; CAA43292.1; JOINED.
 DR EMBL: X61053; CAA43292.1; JOINED.
 DR EMBL: X61054; CAA43292.1; JOINED.
 DR EMBL: X61055; CAA43292.1; JOINED.
 DR EMBL: X61056; CAA43292.1; JOINED.
 DR EMBL: X61057; CAA43292.1; JOINED.
 DR EMBL: X61058; CAA43292.1; JOINED.
 DR EMBL: X61059; CAA43292.1; JOINED.
 DR EMBL: X61060; CAA43292.1; JOINED.
 DR EMBL: X61061; CAA43292.1; JOINED.
 DR EMBL: X61062; CAA43292.1; JOINED.
 DR EMBL: X61063; CAA43292.1; JOINED.
 DR EMBL: X61064; CAA43292.1; JOINED.
 DR EMBL: X61065; CAA43292.1; JOINED.
 DR EMBL: X61066; CAA43292.1; JOINED.
 DR EMBL: X61067; CAA43292.1; JOINED.
 DR EMBL: X61068; CAA43292.1; JOINED.
 DR EMBL: X61069; CAA43292.1; JOINED.
 DR EMBL: X61070; CAA43292.1; JOINED.
 DR EMBL: X61071; CAA43292.1; JOINED.
 DR EMBL: X61072; CAA43292.1; JOINED.
 DR EMBL: X61073; CAA43292.1; JOINED.
 DR EMBL: X61074; CAA43292.1; JOINED.
 DR EMBL: X61075; CAA43292.1; JOINED.
 DR EMBL: X61076; CAA43292.1; JOINED.
 DR EMBL: X61077; CAA43292.1; JOINED.
 DR EMBL: X61078; CAA43292.1; JOINED.
 DR EMBL: X61079; CAA43292.1; JOINED.
 DR EMBL: X61080; CAA43292.1; JOINED.
 DR EMBL: X61081; CAA43292.1; JOINED.
 DR EMBL: X61082; CAA43292.1; JOINED.
 DR EMBL: X61083; CAA43292.1; JOINED.
 DR EMBL: X61084; CAA43292.1; JOINED.
 DR EMBL: X61085; CAA43292.1; JOINED.
 DR EMBL: X61086; CAA43292.1; JOINED.
 DR EMBL: X61087; CAA43292.1; JOINED.
 DR EMBL: X61088; CAA43292.1; JOINED.
 DR EMBL: X61089; CAA43292.1; JOINED.
 DR EMBL: X61090; CAA43292.1; JOINED.
 DR EMBL: X61091; CAA43292.1; JOINED.
 DR EMBL: X61092; CAA43292.1; JOINED.
 DR EMBL: X61093; CAA43292.1; JOINED.
 DR EMBL: X61094; CAA43292.1; JOINED.
 DR EMBL: X61095; CAA43292.1; JOINED.
 DR EMBL: X61096; CAA43292.1; JOINED.
 DR EMBL: X61097; CAA43292.1; JOINED.
 DR EMBL: X61098; CAA43292.1; JOINED.
 DR EMBL: X61099; CAA43292.1; JOINED.
 DR EMBL: X61100; CAA43292.1; JOINED.
 DR EMBL: X61101; CAA43292.1; JOINED.
 DR EMBL: X61102; CAA43292.1; JOINED.
 DR EMBL: X61103; CAA43292.1; JOINED.
 DR EMBL: X61104; CAA43292.1; JOINED.
 DR EMBL: X61105; CAA43292.1; JOINED.
 DR EMBL: X61106; CAA43292.1; JOINED.
 DR EMBL: X61107; CAA43292.1; JOINED.
 DR EMBL: X61108; CAA43292.1; JOINED.
 DR EMBL: X61109; CAA43292.1; JOINED.
 DR EMBL: X61110; CAA43292.1; JOINED.
 DR EMBL: X61111; CAA43292.1; JOINED.
 DR EMBL: X61112; CAA43292.1; JOINED.
 DR EMBL: X61113; CAA43292.1; JOINED.
 DR EMBL: X61114; CAA43292.1; JOINED.
 DR EMBL: X61115; CAA43292.1; JOINED.
 DR EMBL: X61116; CAA43292.1; JOINED.
 DR EMBL: X61117; CAA43292.1; JOINED.
 DR EMBL: X61118; CAA43292.1; JOINED.
 DR EMBL: X61119; CAA43292.1; JOINED.
 DR EMBL: X61120; CAA43292.1; JOINED.
 DR EMBL: X61121; CAA43292.1; JOINED.
 DR EMBL: X61122; CAA43292.1; JOINED.
 DR EMBL: X61123; CAA43292.1; JOINED.
 DR EMBL: X61124; CAA43292.1; JOINED.
 DR EMBL: X61125; CAA43292.1; JOINED.
 DR EMBL: X61126; CAA43292.1; JOINED.
 DR EMBL: X61127; CAA43292.1; JOINED.
 DR EMBL: X61128; CAA43292.1; JOINED.
 DR EMBL: X61129; CAA43292.1; JOINED.
 DR EMBL: X61130; CAA43292.1; JOINED.
 DR EMBL: X61131; CAA43292.1; JOINED.
 DR EMBL: X61132; CAA43292.1; JOINED.
 DR EMBL: X61133; CAA43292.1; JOINED.
 DR EMBL: X61134; CAA43292.1; JOINED.
 DR EMBL: X61135; CAA43292.1; JOINED.
 DR EMBL: X61136; CAA43292.1; JOINED.
 DR EMBL: X61137; CAA43292.1; JOINED.
 DR EMBL: X61138; CAA43292.1; JOINED.
 DR EMBL: X61139; CAA43292.1; JOINED.
 DR EMBL: X61140; CAA43292.1; JOINED.
 DR EMBL: X61141; CAA43292.1; JOINED.
 DR EMBL: X61142; CAA43292.1; JOINED.
 DR EMBL: X61143; CAA43292.1; JOINED.
 DR EMBL: X61144; CAA43292.1; JOINED.
 DR EMBL: X61145; CAA43292.1; JOINED.
 DR EMBL: X61146; CAA43292.1; JOINED.
 DR EMBL: X61147; CAA43292.1; JOINED.
 DR EMBL: X61148; CAA43292.1; JOINED.
 DR EMBL: X61149; CAA43292.1; JOINED.
 DR EMBL: X61150; CAA43292.1; JOINED.
 DR EMBL: X61151; CAA43292.1; JOINED.
 DR EMBL: X61152; CAA43292.1; JOINED.
 DR EMBL: X61153; CAA43292.1; JOINED.
 DR EMBL: X61154; CAA43292.1; JOINED.
 DR EMBL: X61155; CAA43292.1; JOINED.
 DR EMBL: X61156; CAA43292.1; JOINED.
 DR EMBL: X61157; CAA43292.1; JOINED.
 DR EMBL: X61158; CAA43292.1; JOINED.
 DR EMBL: X61159; CAA43292.1; JOINED.
 DR EMBL: X61160; CAA43292.1; JOINED.
 DR EMBL: X61161; CAA43292.1; JOINED.
 DR EMBL: X61162; CAA43292.1; JOINED.
 DR EMBL: X61163; CAA43292.1; JOINED.
 DR EMBL: X61164; CAA43292.1; JOINED.
 DR EMBL: X61165; CAA43292.1; JOINED.
 DR EMBL: X61166; CAA43292.1; JOINED.
 DR EMBL: X61167; CAA43292.1; JOINED.
 DR EMBL: X61168; CAA43292.1; JOINED.
 DR EMBL: X61169; CAA43292.1; JOINED.
 DR EMBL: X61170; CAA43292.1; JOINED.
 DR EMBL: X61171; CAA43292.1; JOINED.
 DR EMBL: X61172; CAA43292.1; JOINED.
 DR EMBL: X61173; CAA43292.1; JOINED.
 DR EMBL: X61174; CAA43292.1; JOINED.
 DR EMBL: X61175; CAA43292.1; JOINED.
 DR EMBL: X61176; CAA43292.1; JOINED.
 DR EMBL: X61177; CAA43292.1; JOINED.
 DR EMBL: X61178; CAA43292.1; JOINED.
 DR EMBL: X61179; CAA43292.1; JOINED.
 DR EMBL: X61180; CAA43292.1; JOINED.
 DR EMBL: X61181; CAA43292.1; JOINED.
 DR EMBL: X61182; CAA43292.1; JOINED.
 DR EMBL: X61183; CAA43292.1; JOINED.
 DR EMBL: X61184; CAA43292.1; JOINED.
 DR EMBL: X61185; CAA43292.1; JOINED.
 DR EMBL: X61186; CAA43292.1; JOINED.
 DR EMBL: X61187; CAA43292.1; JOINED.
 DR EMBL: X61188; CAA43292.1; JOINED.
 DR EMBL: X61189; CAA43292.1; JOINED.
 DR EMBL: X61190; CAA43292.1; JOINED.
 DR EMBL: X61191; CAA43292.1; JOINED.
 DR EMBL: X61192; CAA43292.1; JOINED.
 DR EMBL: X61193; CAA43292.1; JOINED.
 DR EMBL: X61194; CAA43292.1; JOINED.
 DR EMBL: X61195; CAA43292.1; JOINED.
 DR EMBL: X61196; CAA43292.1; JOINED.
 DR EMBL: X61197; CAA43292.1; JOINED.
 DR EMBL: X61198; CAA43292.1; JOINED.
 DR EMBL: X61199; CAA43292.1; JOINED.
 DR EMBL: X61200; CAA43292.1; JOINED.
 DR EMBL: X61201; CAA43292.1; JOINED.
 DR EMBL: X61202; CAA43292.1; JOINED.
 DR EMBL: X61203; CAA43292.1; JOINED.
 DR EMBL: X61204; CAA43292.1; JOINED.
 DR EMBL: X61205; CAA43292.1; JOINED.
 DR EMBL: X61206; CAA43292.1; JOINED.
 DR EMBL: X61207; CAA43292.1; JOINED.
 DR EMBL: X61208; CAA43292.1; JOINED.
 DR EMBL: X61209; CAA43292.1; JOINED.
 DR EMBL: X61210; CAA43292.1; JOINED.
 DR EMBL: X61211; CAA43292.1; JOINED.
 DR EMBL: X61212; CAA43292.1; JOINED.
 DR EMBL: X61213; CAA43292.1; JOINED.
 DR EMBL: X61214; CAA43292.1; JOINED.
 DR EMBL: X61215; CAA43292.1; JOINED.
 DR EMBL: X61216; CAA43292.1; JOINED.
 DR EMBL: X61217; CAA43292.1; JOINED.
 DR EMBL: X61218; CAA43292.1; JOINED.
 DR EMBL: X61219; CAA43292.1; JOINED.
 DR EMBL: X61220; CAA43292.1; JOINED.
 DR EMBL: X61221; CAA43292.1; JOINED.
 DR EMBL: X61222; CAA43292.1; JOINED.
 DR EMBL: X61223; CAA43292.1; JOINED.
 DR EMBL: X61224; CAA43292.1; JOINED.
 DR EMBL: X61225; CAA43292.1; JOINED.
 DR EMBL: X61226; CAA43292.1; JOINED.
 DR EMBL: X61227; CAA43292.1; JOINED.
 DR EMBL: X61228; CAA43292.1; JOINED.
 DR EMBL: X61229; CAA43292.1; JOINED.
 DR EMBL: X61230; CAA43292.1; JOINED.
 DR EMBL: X61231; CAA43292.1; JOINED.
 DR EMBL: X61232; CAA43292.1; JOINED.
 DR EMBL: X61233; CAA43292.1; JOINED.
 DR EMBL: X61234; CAA43292.1; JOINED.
 DR EMBL: X61235; CAA43292.1; JOINED.
 DR EMBL: X61236; CAA43292.1; JOINED.
 DR EMBL: X61237; CAA43292.1; JOINED.
 DR EMBL: X61238; CAA43292.1; JOINED.
 DR EMBL: X61239; CAA43292.1; JOINED.
 DR EMBL: X61240; CAA43292.1; JOINED.
 DR EMBL: X61241; CAA43292.1; JOINED.
 DR EMBL: X61242; CAA43292.1; JOINED.
 DR EMBL: X61243; CAA43292.1; JOINED.
 DR EMBL: X61244; CAA43292.1; JOINED.
 DR EMBL: X61245; CAA43292.1; JOINED.
 DR EMBL: X61246; CAA43292.1; JOINED.
 DR EMBL: X61247; CAA43292.1; JOINED.
 DR EMBL: X61248; CAA43292.1; JOINED.
 DR EMBL: X61249; CAA43292.1; JOINED.
 DR EMBL: X61250; CAA43292.1; JOINED.
 DR EMBL: X61251; CAA43292.1; JOINED.
 DR EMBL: X61252; CAA43292.1; JOINED.
 DR EMBL: X61253; CAA43292.1; JOINED.
 DR EMBL: X61254; CAA43292.1; JOINED.
 DR EMBL: X61255; CAA43292.1; JOINED.
 DR EMBL: X61256; CAA43292.1; JOINED.
 DR EMBL: X61257; CAA43292.1; JOINED.
 DR EMBL: X61258; CAA43292.1; JOINED.
 DR EMBL: X61259; CAA43292.1; JOINED.
 DR EMBL: X61260; CAA43292.1; JOINED.
 DR EMBL: X61261; CAA43292.1; JOINED.
 DR EMBL: X61262; CAA43292.1; JOINED.
 DR EMBL: X61263; CAA43292.1; JOINED.
 DR EMBL: X61264; CAA43292.1; JOINED.
 DR EMBL: X61265; CAA43292.1; JOINED.
 DR EMBL: X61266; CAA43292.1; JOINED.
 DR EMBL: X61267; CAA43292.1; JOINED.
 DR EMBL: X61268; CAA43292.1; JOINED.
 DR EMBL: X61269; CAA43292.1; JOINED.
 DR EMBL: X61270; CAA43292.1; JOINED.
 DR EMBL: X61271; CAA43292.1; JOINED.
 DR EMBL: X61272; CAA43292.1; JOINED.
 DR EMBL: X61273; CAA43292.1; JOINED.
 DR EMBL: X61274; CAA43292.1; JOINED.
 DR EMBL: X61275; CAA43292.1; JOINED.
 DR EMBL: X61276; CAA43292.1; JOINED.
 DR EMBL: X61277; CAA43292.1; JOINED.
 DR EMBL: X61278; CAA43292.1; JOINED.
 DR EMBL: X61279; CAA43292.1; JOINED.
 DR EMBL: X61280; CAA43292.1; JOINED.
 DR EMBL: X61281; CAA43292.1; JOINED.
 DR EMBL: X61282; CAA43292.1; JOINED.
 DR EMBL: X61283; CAA43292.1; JOINED.
 DR EMBL: X61284; CAA43292.1; JOINED.
 DR EMBL: X61285; CAA43292.1; JOINED.
 DR EMBL: X61286; CAA43292.1; JOINED.
 DR EMBL: X61287; CAA43292.1; JOINED.
 DR EMBL: X61288; CAA43292.1; JOINED.
 DR EMBL: X61289; CAA43292.1; JOINED.
 DR EMBL: X61290; CAA43292.1; JOINED.
 DR EMBL: X61291; CAA43292.1; JOINED.
 DR EMBL: X61292; CAA43292.1; JOINED.
 DR EMBL: X61293; CAA43292.1; JOINED.
 DR EMBL: X61294; CAA43292.1; JOINED.
 DR EMBL: X61295; CAA43292.1; JOINED.
 DR EMBL: X61296; CAA43292.1; JOINED.
 DR EMBL: X61297; CAA43292.1; JOINED.
 DR EMBL: X61298; CAA43292.1; JOINED.
 DR EMBL: X61299; CAA43292.1; JOINED.
 DR EMBL: X61300; CAA43292.1; JOINED.
 DR EMBL: X61301; CAA43292.1; JOINED.
 DR EMBL: X61302; CAA43292.1; JOINED.
 DR EMBL: X61303; CAA43292.1; JOINED.
 DR EMBL: X61304; CAA43292.1; JOINED.
 DR EMBL: X61305; CAA43292.1; JOINED.
 DR EMBL: X61306; CAA43292.1; JOINED.
 DR EMBL: X61307; CAA43292.1; JOINED.
 DR EMBL: X61308; CAA43292.1; JOINED.
 DR EMBL: X61309; CAA43292.1; JOINED.
 DR EMBL: X61310; CAA43292.1; JOINED.
 DR EMBL: X61311; CAA43292.1; JOINED.
 DR EMBL: X61312; CAA43292.1; JOINED.
 DR EMBL: X61313; CAA43292.1; JOINED.
 DR EMBL: X61314; CAA43292.1; JOINED.
 DR EMBL: X61315; CAA43292.1; JOINED.
 DR EMBL: X61316; CAA43292.1; JOINED.
 DR EMBL: X61317; CAA43292.1; JOINED.
 DR EMBL: X61318; CAA43292.1; JOINED.
 DR EMBL: X61319; CAA43292.1; JOINED.
 DR EMBL: X61320; CAA43292.1; JOINED.
 DR EMBL: X61321; CAA43292.1; JOINED.
 DR EMBL: X61322; CAA43292.1; JOINED.
 DR EMBL: X61323; CAA43292.1; JOINED.
 DR EMBL: X61324; CAA43292.1; JOINED.
 DR EMBL: X61325; CAA43292.1; JOINED.
 DR EMBL: X61326; CAA43292.1; JOINED.
 DR EMBL: X61327; CAA43292.1; JOINED.
 DR EMBL: X61328; CAA43292.1; JOINED.
 DR EMBL: X61329; CAA43292.1; JOINED.
 DR EMBL: X61330; CAA43292.1; JOINED.
 DR EMBL: X61331; CAA43292.1; JOINED.
 DR EMBL: X61332; CAA43292.1; JOINED.

```

RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92363956.
RA Salomon D., Ayala O., Patel-King R., Hynes R.O., Geiger B.:
RT "Extrajunctional distribution of N-cadherin in cultured human
   endothelial cells."
RL J. Cell Sci. 102:7-17(1992).
RN [4]
RP SEQUENCE OF 160-906 FROM N.A.
RX MEDLINE: 90347462.
RA Walsh F.S., Barton C.H., Putt W., Moore S.E., Kelsell D.,
RA Spurr N., Goodfellow P.N.:
RT "N-cadherin gene maps to human chromosome 18 and is not linked to the
   E-cadherin gene."
RL J. Neurochem. 55:805-812(1990).
RN [5]
RP SEQUENCE OF 1-20 FROM N.A.
RX MEDLINE: 95048366.
RA Wallis J.A., Fox M., Walsh F.S.:
RT "Structure of the human N-cadherin gene: VAC analysis and fine
   chromosomal mapping to 18q11.2."
RL Genomics 22:172-179(1994).
CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC NEURONAL RECOGNITION MECHANISM.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X57548; CAA40773.1; -
DR EMBL: X54315; CAA38213.1; -
DR EMBL: S43303; AAB22854.1; -
DR EMBL: M34064; AAA03336.1; -
DR EMBL: Z27420; CAA81799.1; -
DR PIR: A38870; IJHUCN.
DR HSSP: P15116; INCH.
DR MIM: 114020.
DR PRINTS: PR00205; CADHERIN.
DR PROSITE: PS00232; CADHERIN. 3.
DR PFAM: PF01049; Cadherin_C-term; 1.
KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW Calcium-binding; Repeat; Signal.
FT SIGNAL 1 23
FT PROPEP 24 159
FT CHAIN 160 906
FT DOMAIN 160 724
FT TRANSMEM 725 746
FT DOMAIN 747 906
FT REPEAT 160 267
FT REPEAT 268 382
FT REPEAT 383 497
FT REPEAT 498 603
FT REPEAT 604 714
FT DOMAIN 714 878
FT CARBOHYD 190 190
FT CARBOHYD 273 273
FT CARBOHYD 325 325
FT CARBOHYD 402 402
FT CARBOHYD 572 572
FT CARBOHYD 651 651
FT CARBOHYD 692 692
FT CONFLICT 12 12
FT CONFLICT 16 16
L -> A (IN REF. 3 AND 5).

```

```

FT CONFLICT 196 196 S -> T (IN REF. 1).
FT CONFLICT 212 212 I -> L (IN REF. 4).
FT CONFLICT 357 357 N -> I (IN REF. 1).
FT CONFLICT 867 867 A -> L (IN REF. 3).
SQ SEQUENCE 906 AA; 99851 MW; 72DDC7BB857C7AFC CRC64;

Query Match 62.5%; Score 60; DB 1; Length 906;
Best Local Similarity 63.6%; Pred. No. 1.30e-01;
Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

DB 300 LNSKIAFRIVSOEPA 314
Oy 1 LNSKIAFRIVSOEPA 15

RESULT 7
ID YOX1_YEAST STANDARD: PRT; 385 AA.
AC P34161;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HOMEBOX PROTEIN YOX1.
GN YOX1 OR YML027W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE OF 1-296 FROM N.A.
RX MEDLINE: 93209080.
RA Kaufmann E.;
RT "In vitro binding to the leucine tRNA gene identifies a novel yeast
   homeobox gene."
RL Chromosoma 102:174-179(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IN VITRO, IS CAPABLE OF BINDING TO THE DNA OF THE
CC LEUCINE tRNA GENE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: WITH OTHER HOMEBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62392; CAA44264.1; -
DR EMBL: Z46659; CAA86628.1; -
DR PIR: S33388; S33388.
DR HSSP: P06601; IFJL.
DR SGD: L0002540; YOX1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
DR PFAM: PF00046; homeobox; 1.
KW Homeobox; DNA-binding; Nuclear protein.
FT DNA_BIND 176 235
FT CONFLICT 291 296
FT CONFLICT 296 296
SQ SEQUENCE 385 AA; 42739 MW; F21A95CDD97A282 CRC64;

Query Match 61.5%; Score 59; DB 1; Length 385;
Best Local Similarity 63.6%; Pred. No. 1.30e-01;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 364 KEGKIVDOOP 374
Oy 4 KIAKIVSOEP 14

```

```

RESULT      8
ID          YGN1_YEAST      STANDARD:      PRT: 1264 AA.
AC          P53125:
DC          01-OCT-1996 (Rel. 34, Created)
DT          01-OCT-1996 (Rel. 34, Last sequence update)
DT          15-DEC-1998 (Rel. 37, Last annotation update)
DE          HYPOTHETICAL 145.6 KD PROTEIN IN RPLB-CG1 INTERGENIC REGION.
GN          YGL13W OR G2842.
OS          Saccharomyces cerevisiae (Baker's yeast).
OC          Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC          Saccharomycetaceae; Saccharomyces.
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN-S288C / FY1679;
RA          MEDLINE: 96437978.
RX          Escribano V., Erasó P., Portillo F., Mazon M.J.;
RT          *Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces
RT          cerevisiae chromosome VII reveals SEC27, SSM1b, a putative
RT          S-adenosylmethionine-dependent enzyme and six new open reading
RT          frames.
RL          Yeast 12:887-892(1996).
CC          -----
CC          This SWISS-PROT entry is copyright. It is produced through a collaboration
CC          between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC          the European Bioinformatics Institute. There are no restrictions on its
CC          use by non-profit institutions as long as its content is in no way
CC          modified and this statement is not removed. Usage by and for commercial
CC          entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC          or send an email to license@sib-sib.ch).
CC          -----
DR          EMBL: 272655; CAA96844.1;
DR          Hypothetical protein.
SQ          SEQUENCE 1264 AA: 145642 MW: 45E4CF8835C7C746 CRC64;

Query Match      59.48; Score 57; DB 1; Length 1264;
deest Local Similarity 42.9%; Pred. No. 3,79e-01;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db      524 SDKVEKIVDDPA 537
QY      2 NSKIARIVSOEPA 15

RESULT      9
ID          FAT_DROME      STANDARD:      PRT: 5147 AA.
AC          P33450:
DC          01-FEB-1994 (Rel. 28, Created)
DT          01-FEB-1996 (Rel. 33, Last sequence update)
DT          15-JUL-1998 (Rel. 36, Last annotation update)
DE          CADHERIN-RELATED TUMOR SUPPRESSOR PRECURSOR (FAT PROTEIN).
GN          FT.
OS          Drosophila melanogaster (Fruit fly).
OC          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
CC          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CC          Ephydriidae; Drosophilidae; Drosophila.
RN          [1]
RP          SEQUENCE FROM N.A.
RX          MEDLINE: 92069752.
RA          Mahoney P.A., Weber U., Onofrechuk P., Blessmann H., Bryant P.J.,
RA          Goodman C.S.;
RT          *The fat tumor suppressor gene in Drosophila encodes a novel member
RT          of the cadherin gene superfamily.
RL          Cell 67:853-868(1991).
CC          -1- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN.
CC          -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC          -1- DISEASE: RECESSIVE LETHAL MUTATIONS IN FAT CAUSE HYPERPLASTIC,
CC          TUMOR-LIKE OVERGROWTH OF LARVAL IMAGINAL DISCS, DEFECTS IN
CC          DIFFERENTIATION AND MORPHOGENESIS, AND DEATH DURING THE PUPAL
CC          STAGE.
CC          -1- MISCELLANEOUS: SIMILARITY: CONTAINS 37 CADHERINS-TYPE REPEATS.
CC          -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC          -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC          -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.

```

```

CC          -----
CC          This SWISS-PROT entry is copyright. It is produced through a collaboration
CC          between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC          the European Bioinformatics Institute. There are no restrictions on its
CC          use by non-profit institutions as long as its content is in no way
CC          modified and this statement is not removed. Usage by and for commercial
CC          entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC          or send an email to license@sib-sib.ch).
CC          -----
DR          EMBL: M80537; AAA28530.1;
DR          PIR: A41087; IJFPTM.
DR          HSSP: P00740; IIXA.
DR          FLYBASE: FBgn0001075; ft.
DR          PRINTS: PR00205; CADHERIN.
DR          PROSITE: PS00232; CADHERIN_22.
DR          PROSITE: PS00022; EGF_1; 4.
DR          PROSITE: PS01186; EGF_2; 2.
DR          PFM: PF00008; EGF; 4.
DR          PFM: PF00028; cadherin; 34.
DR          PFM: PF00054; laminin_G; 2.
KM          Cell adhesion. Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KM          Calcium-binding; Repeat; EGF-like domain.
KW          SIGNAL 1 35
FT          CHAIN 36 5147
FT          DOMAIN 36 4583
FT          TRANSMEM 4584 4609
FT          DOMAIN 4610 5147
FT          REPEAT 36 156
FT          REPEAT 157 270
FT          REPEAT 271 382
FT          REPEAT 383 494
FT          REPEAT 495 599
FT          REPEAT 600 708
FT          REPEAT 709 821
FT          REPEAT 821 942
FT          REPEAT 943 1049
FT          REPEAT 1050 1153
FT          REPEAT 1154 1278
FT          REPEAT 1279 1384
FT          REPEAT 1385 1489
FT          REPEAT 1490 1601
FT          REPEAT 1602 1713
FT          REPEAT 1714 1823
FT          REPEAT 1824 1922
FT          REPEAT 1923 2027
FT          REPEAT 2028 2167
FT          REPEAT 2168 2278
FT          REPEAT 2279 2385
FT          REPEAT 2386 2491
FT          REPEAT 2492 2596
FT          REPEAT 2597 2703
FT          REPEAT 2704 2810
FT          REPEAT 2811 2913
FT          REPEAT 2914 3013
FT          REPEAT 3014 3124
FT          REPEAT 3125 3229
FT          REPEAT 3230 3334
FT          REPEAT 3335 3439
FT          REPEAT 3440 3545
FT          REPEAT 3546 3651
FT          REPEAT 3652 3756
FT          REPEAT 3757 4011
FT          DOMAIN 4013 4049
FT          DOMAIN 4052 4090
FT          DOMAIN 4092 4128
FT          DOMAIN 4321 4362
FT          DISULFID 3954 3966
FT          DISULFID 3960 3999
FT          DISULFID 4001 4010
FT          DISULFID 4017 4028
FT          DISULFID 4022 4037
FT          DISULFID 4039 4048
FT          DISULFID 4056 4067

```

```

FT DISULFID 4061 4078 BY SIMILARITY.
FT DISULFID 4080 4089 BY SIMILARITY.
FT DISULFID 4096 4107 BY SIMILARITY.
FT DISULFID 4101 4116 BY SIMILARITY.
FT DISULFID 4118 4127 BY SIMILARITY.
FT DISULFID 4325 4341 BY SIMILARITY.
FT DISULFID 4334 4350 BY SIMILARITY.
FT DISULFID 4352 4361 BY SIMILARITY.
FT CARBOHYD 239 239 POTENTIAL.
FT CARBOHYD 257 257 POTENTIAL.
FT CARBOHYD 276 276 POTENTIAL.
FT CARBOHYD 280 280 POTENTIAL.
FT CARBOHYD 402 402 POTENTIAL.
FT CARBOHYD 461 461 POTENTIAL.
FT CARBOHYD 605 605 POTENTIAL.
FT CARBOHYD 631 631 POTENTIAL.
FT CARBOHYD 1155 1155 POTENTIAL.
FT CARBOHYD 1367 1367 POTENTIAL.
FT CARBOHYD 1458 1458 POTENTIAL.
FT CARBOHYD 1751 1751 POTENTIAL.
FT CARBOHYD 1831 1831 POTENTIAL.
FT CARBOHYD 1880 1880 POTENTIAL.
FT CARBOHYD 2080 2080 POTENTIAL.
FT CARBOHYD 2171 2171 POTENTIAL.
FT CARBOHYD 2247 2247 POTENTIAL.
FT CARBOHYD 2290 2290 POTENTIAL.
FT CARBOHYD 2437 2437 POTENTIAL.
FT CARBOHYD 2581 2581 POTENTIAL.
FT CARBOHYD 2799 2799 POTENTIAL.
FT CARBOHYD 2920 2920 POTENTIAL.
FT CARBOHYD 2946 2946 POTENTIAL.
FT CARBOHYD 2967 2967 POTENTIAL.
FT CARBOHYD 3157 3157 POTENTIAL.
FT CARBOHYD 3303 3303 POTENTIAL.
FT CARBOHYD 3386 3386 POTENTIAL.
FT CARBOHYD 3389 3389 POTENTIAL.
FT CARBOHYD 3525 3525 POTENTIAL.
FT CARBOHYD 3852 3852 POTENTIAL.
FT CARBOHYD 3865 3865 POTENTIAL.
FT CARBOHYD 3905 3905 POTENTIAL.
FT CARBOHYD 4306 4306 POTENTIAL.
FT CARBOHYD 4414 4414 POTENTIAL.
FT CARBOHYD 4471 4471 POTENTIAL.
FT CARBOHYD 4487 4487 POTENTIAL.
FT CARBOHYD 4539 4539 POTENTIAL.
FT CARBOHYD 4550 4550 POTENTIAL.
FT VARIANT 1229 1229 S -> G.
FT VARIANT 1233 1233 G -> S.
SQ SEQUENCE 5147 AA; 564868 MW; 4BE85ED6C12397EE CRC64;

Query Match 59.4%; Score 57; DB 1; Length 5147;
Best Local Similarity 50.0%; Pred. No. 3.79e-01;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

```

Db 1520 LNSKIATKIVSOEP 1533
Oy 1 LNSKIATKIVSOEP 14

```

```

RESULT 10
ID EFP_BORBU STANDARD: PRT; 192 AA.
AC 051232;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE ELONGATION FACTOR P (EF-P).
GN EFP OR BB0214.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 35210 / B31;
RX MEDLINE: 98065943.

```

```

RA Fraser C.M., Castjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kierlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Uitterback T., Matthey L., McDonald L., Atliach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi."
RL Nature 390:580-586(1997).
CC -1- FUNCTION: INVOLVED IN PEPTIDE BOND SYNTHESIS. STIMULATE EFFICIENT
CC TRANSLATION AND PEPTIDE-BOND SYNTHESIS ON NATIVE OR RECONSTITUTED
CC 70S RIBOSOMES IN VITRO. PROBABLY FUNCTIONS INDIRECTLY BY ALTERING
CC THE AFFINITY OF THE RIBOSOME FOR AMINOACYL-TRNA, THUS INCREASING
CC THEIR REACTIVITY AS ACCEPTORS FOR PEPTIDYL TRANSFERASE
CC (BY SIMILARITY).
CC -1- PATHWAY: PROTEIN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE ELONGATION FACTOR P FAMILY.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC .....
CC DR EMBL: AE001132; AAC66610.1; -.
CC TIGR: BB0214; -.
DR PROSITE: PS01275; EFP; FALSE_NEG.
DR PFAM: PF01132; EFP; 1.
KW Protein biosynthesis; Elongation factor.
SQ SEQUENCE 192 AA; 21413 MW; AC2C48A8BC73DE0 CRC64;

```

```

Query Match 57.3%; Score 55; DB 1; Length 192;
Best Local Similarity 53.3%; Pred. No. 1.08e-00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

Db 130 LARKIPEVEVEEAA 144
Oy 1 LNSKIATKIVSOEPA 15

```

```

RESULT 11
ID CAD2_BOVIN STANDARD: PRT; 877 AA.
AC P19534;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NEURAL-CADHERIN PRECURSOR (N-CADHERIN) (FRAGMENT).
GN CDR2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90360979.
RA Liaw C.W., Cannon C., Power M.D., Kibonoka P.K., Rubin L.L.;
RT "Identification and cloning of two species of cadherins in bovine
RT endothelial cells."
RL EMBL J. 9:2701-2708(1990).
CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC NEURONAL RECOGNITION MECHANISM.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).

DR EMBL: X53615: CAA37677.1: -
 DR PIR: S11693: IBOCN.
 DR HSP: P15116: INCH.
 DR PROSITE: PS00232: CADHERIN: 3.
 DR PFAM: PF00028: cadherin: 5.
 DR PFAM: PF01049: Cadherin_C-term: 1.
 KM Cell adhesion: Glycoprotein: Phosphorylation: Transmembrane:
 KM Calcium-binding: Repeat.
 FT NON-TER 1
 FT PROPEP <1 130
 FT CHAIN 131 877 NEURAL-CADHERIN.
 FT DOMAIN 131 695 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 696 717 POTENTIAL.
 FT DOMAIN 718 877 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 131 238 CADHERIN 1.
 FT REPEAT 239 363 CADHERIN 2.
 FT REPEAT 364 468 CADHERIN 3.
 FT REPEAT 469 574 CADHERIN 4.
 FT REPEAT 575 685 CADHERIN 5.
 FT DOMAIN 834 849 SER-RICH.
 FT CARBOHYD 161 161 POTENTIAL.
 FT CARBOHYD 244 244 POTENTIAL.
 FT CARBOHYD 296 296 POTENTIAL.
 FT CARBOHYD 373 373 POTENTIAL.
 FT CARBOHYD 543 543 POTENTIAL.
 FT CARBOHYD 622 622 POTENTIAL.
 FT CARBOHYD 663 663 POTENTIAL.
 SQ SEQUENCE 877 AA: 96845 MW: 44188295ED871A249 CRC64:
 Query Match 56.3% Score 54: DB 1: Length 877:
 Best Local Similarity 40.0% Pred. No. 1.79e+00:
 Matches 6: Conservative 7: Mismatches 2: Indels 0: Gaps 0:

Db 271 LNCMR:R1LSOAPS 285
 1 LNSKIAFKIVSOEPA 15

RESULT 12
 ID CADI_HUMAN STANDARD: PRT: 882 AA.
 AC P12830: Q14216:
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE EPITHELIAL-CADHERIN PRECURSOR (E-CADHERIN) (UWOMORULIN) (CAM 120/80).
 GN CDH1 OR UVO OR CDHE.
 OS Homo sapiens (Human).
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
 OC Eutheria: Primates: Catarrhini: Hominoidea: Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92311394.
 RA Bussemakers M.J.G., Mees S.G.M., van Bokhoven A., Debruyne F.M.J.,
 RA Schalken J.A.:
 RT "Molecular cloning and characterization of the human E-cadherin
 RT CDNA.";
 RL Mol. Biol. Rep. 17:123-128(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kelder W., Wada A., Oda T., Hirohashi S., Kemler R., Birchmeier W.:
 RL Submitted (xxx-1992) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 172-311 FROM N.A.
 RX TISSUE-LIVER:
 RX MEDLINE: 89031725.
 RA Mansouri A., Spurr N., Goodfellow P.N., Kemler R.:
 RT "Characterization and chromosomal localization of the gene encoding

RT the human cell adhesion molecule uvomorulin.";
 RL Differentiation 38:67-71(1988).
 RN [4]
 RP SEQUENCE OF 1-16 FROM N.A.
 RX MEDLINE: 94380041.
 RA Bussemakers M.J., Giroldi L.A., van Bokhoven A., Schalken J.A.:
 RT "Transcriptional regulation of the human E-cadherin gene in human
 RT prostate cancer cell lines: Characterization of the human E-cadherin
 RT gene promoter.";
 RL Biochem. Biophys. Res. Commun. 203:1284-1290(1994).
 RN [5]
 RP REVIEW ON VARIANTS.
 RX MEDLINE: 98415721.
 RA Berr G., Becker K.-F., Hoefler H., van Roy F.:
 RT "Mutations of the human E-cadherin (CDH1) gene.";
 RL Hum. Mutat. 12:226-237(1998).
 RN [6]
 RP VARIANTS ALA-370 AND ASN-473.
 RX MEDLINE: 94306394.
 RA Becker K.-F., Atkinson M.J., Reich U., Becker I., Nekarda H.,
 RT Stewart J.R., Hoefler H.:
 RL "E-cadherin gene mutations provide clues to diffuse type gastric
 RL carcinomas.";
 RN Cancer Res. 54:3845-3852(1994).
 RN [7]
 RP VARIANT LOBULAR BREAST CARCINOMA SER-315.
 RX MEDLINE: 95049851.
 RA Kanal Y., Oda T., Tsuda H., Ochiai A., Hirohashi S.:
 RT "Point mutation of the E-cadherin gene in invasive lobular carcinoma
 RT of the breast.";
 RL Jpn. J. Cancer Res. 85:1035-1039(1994).
 RN [8]
 RP VARIANTS GYNECOLOGIC CANCERS THR-617; VAL-711 AND GLY-838.
 RX MEDLINE: 94353983.
 RA Risinger J.I., Berchuck A., Kohler M.F., Boyd J.:
 RT "Mutations of the E-cadherin gene in human gynecologic cancers.";
 RL Nat. Genet. 7:98-102(1994).
 RN [9]
 RP VARIANT GASTRIC ADENOCARCINOMA GLY-274--PRO-277 DEL.
 RX MEDLINE: 94173928.
 RA Oda T., Kanal Y., Oyama T., Yoshitura K., Shimoyama Y., Birchmeier W.,
 RA Sugimura T., Hirohashi S.:
 RT "E-cadherin gene mutations in human gastric carcinoma cell lines.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:1858-1862(1994).
 RN [10]
 RP VARIANT GASTRIC CARCINOMA PRO-193.
 RX MEDLINE: 96390918.
 RA Muta H., Noguchi M., Kanal Y., Ochiai A., Nawata H., Hirohashi S.:
 RT "E-cadherin gene mutations in signet ring cell carcinoma of the
 RT stomach.";
 RL Jpn. J. Cancer Res. 87:843-848(1996).
 RN [11]
 RP VARIANTS GASTRIC CARCINOMA ASP-400 DEL AND ASP-418--PHE-423 DEL.
 RX MEDLINE: 97197648.
 RA Tamura G., Sakata K., Nishizuka S., Maesawa C., Suzuki Y., Iwaya T.,
 RA Terashima M., Saito K., Satodate R.:
 RT "Inactivation of the E-cadherin gene in primary gastric carcinomas and
 RT gastric carcinoma cell lines.";
 RL Jpn. J. Cancer Res. 87:1153-1159(1996).
 RN [12]
 RP VARIANT THYROID TUMOR THR-592.
 RX MEDLINE: 97138061.
 RA Soares P., Berr G., van Roy F., Sobrinho-Simoes M.:
 RT "E-cadherin gene alterations are rare events in thyroid tumors.";
 RL Int. J. Cancer 70:32-38(1997).
 RN [13]
 RP VARIANTS ASP-336 AND ILE-470.
 RX MEDLINE: 98196671.
 RA Guilford P., Hopkins J., Harraway J., McLeod M., McLeod N.,
 RA Harawira P., Taitte H., Scoular R., Miller A., Reeve A.E.:
 RT "E-cadherin germline mutations in familial gastric cancer.";
 RL Nature 392:402-405(1998).
 CC -I- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.

| | |
|----|---|
| CC | MANNER IN CONNECTING CELLS: CADHERINS MAY THUS CONTRIBUTE TO THE |
| CC | SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN |
| CC | NEURONAL RECOGNITION MECHANISM. |
| CC | -1 SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. |
| CC | -1 SIMILARITY: BELONGS TO THE CADHERIN FAMILY. |
| CC | ----- |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration |
| CC | between the Swiss Institute of Bioinformatics and the EMBL Outstation - |
| CC | the European Bioinformatics Institute. There are no restrictions on its |
| CC | use by non-profit institutions as long as its content is in no way |
| CC | modified and this statement is not removed. Usage by and for commercial |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/license/). |
| CC | or send an email to license@isb-sib.ch . |
| CC | ----- |
| DR | EMBL; M31131; AAA37353.1; .. |
| DR | EMBL; AB008811; BAA23549.1; .. |
| DR | PIR; A32759; IJMSCN. |
| DR | PDB; 1NCG; 10-JUL-95. |
| DR | PDB; 1NCH; 10-JUL-95. |
| DR | PDB; 1NCT; 10-JUL-95. |
| DR | PDB; 1NCU; 18-MAR-99. |
| DR | MGI; 88355; CDH2. |
| DR | PRINTS; PR00205; CADHERIN. |
| DR | PROSITE; PS00232; CADHERIN; 3. |
| DR | PFAM; PF00028; cadherin; 5. |
| DR | PFAM; PF01049; cadherin_C_term; 1. |
| KM | Cell adhesion: Glycoprotein; Phosphorylation; Transmembrane; |
| KM | Calcium-binding; Repeat; Signal; 3D-structure. |
| FT | Cadherin-binding; Repeat; Signal; 3D-structure. |
| FT | PROPEP 1 23 |
| FT | CHAIN 24 159 |
| FT | TRANSMEM 160 906 |
| FT | DOMAIN 160 724 |
| FT | TRANSMEM 725 746 |
| FT | DOMAIN 747 906 |
| FT | REPEAT 160 267 |
| FT | REPEAT 268 382 |
| FT | REPEAT 383 497 |
| FT | REPEAT 498 603 |
| FT | REPEAT 604 717 |
| FT | DOMAIN 863 878 |
| FT | CARBOHYD 190 190 |
| FT | CARBOHYD 273 273 |
| FT | CARBOHYD 325 325 |
| FT | CARBOHYD 402 402 |
| FT | CARBOHYD 572 572 |
| FT | CARBOHYD 651 651 |
| FT | CARBOHYD 692 692 |
| FT | CONFLICT 7 9 |
| FT | CONFLICT 565 565 |
| FT | CONFLICT 567 567 |
| FT | CONFLICT 624 624 |
| SO | SEQUENCE 906 AA; 99761 MW; 0206741B71707B7 CRC64; |
| Dn | Query Match 56.3%; Score 54; DB 1; Length 906; |
| Oy | Best Local Similarity 40.0%; Pred. No. 1.79e+00; |
| | Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0; |
| | 300 LNSKIAFRIVSQEPA 15 |
| | :: : |
| ID | ASIO_YEAST STANDARD; PRT; 1146 AA. |
| AC | P48361; |
| DT | 01-FEB-1996 (Rel. 33, Created) |
| DT | 01-OCT-1996 (Rel. 34, Last sequence update) |
| DE | 01-OCT-1996 (Rel. 34, Last annotation update) |
| | ASK10 PROTEIN. |
| OS | ASK10 OR YGR097W. |
| OS | Saccharomyces cerevisiae (Baker's Yeast). |
| CC | Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales; |
| CC | Saccharomycetaceae; Saccharomyces |

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / SEY6210:
RX MEDLINE: 97060018.
RA Page N., Sheraton J., Brown J.L., Stewart R.S., Bussey H.:
RT "Identification of ASK10 as a multicopy activator of Skn/P-dependent
RL transcription of a HIS3 reporter gene."
RL Yeast 12:267-272(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Hernandez K., Weber N., Wipfl P., Schmidheini T.:
RL Submitted (May-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PUTATIVE ACTIVATOR OF SKN7.
CC -1- SIMILARITY: TO YEAST YIL105C AND YNL047C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U27209; AAA67368.1; -
DR EMBL: Z72882; CA97100.1; -
DR SGD: L0002770; ASK10.
FT DOMAIN 22 26 POLY-GLY.
FT DOMAIN 625 628 POLY-SER.
FT DOMAIN 933 938 POLY-ASN.
FT DOMAIN 958 961 POLY-GLN.
FT DOMAIN 972 975 POLY-SER.
FT CONFLICT 57 57 I -> T (IN REF. 1).
FT CONFLICT 346 346 T -> N (IN REF. 1).
FT CONFLICT 464 464 P -> R (IN REF. 1).
FT CONFLICT 467 467 V -> A (IN REF. 1).
FT CONFLICT 603 603 A -> V (IN REF. 1).
FT CONFLICT 906 906 L -> P (IN REF. 1).
SQ SEQUENCE 1146 AA: 126863 MW: 9E9EF22B0EE496EE CRC64:

Query Match 56.3%; Score 54; DB 1; Length 1146;
Best Local Similarity 70.0%; Pred. No. 1.79e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 692 TRKIVSPEPS 701
QY :||||| 11:
6 AFKIVSOEPA 15

Search completed: Sat May 13 07:07:11 2000
Job time : 90 secs.
```


THIS PAGE BLANK (uspto)

MUSCLE
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Muscle - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 07:07:31 2000; Maspar time 140.71 Seconds

Tabular output not generated. 7.391 Million cell updates/sec

Title: >US-08-991-628-3
Description: (1-15) from US08991628.pep
Perfect Score: 96
Sequence: 1 LNSKIAFKIVSOEPA 15

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 25.549; Variance 30.093; scale 0.849

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|------------|-------|-------------|---------|--------|------------------------|-----------|
| 1 | 92 | 95.8 | 993 11 | 035902 | DESMOGLEIN 3 (FRAGMENT | 2.67e-09 |
| 2 | 61 | 63.5 | 1421 10 | 080807 | T19C21.7 PROTEIN. | 1.77e-01 |
| 3 | 58 | 60.4 | 138 2 | P72670 | HYPOTHETICAL 16.6 KD P | 8.20e-01 |
| 4 | 55 | 57.3 | 225 5 | Q20152 | F38B7.4 PROTEIN. | 3.62e+00 |
| 5 | 55 | 57.3 | 252 14 | Q9YWS9 | ORF MSY032 HYPOTHETICA | 3.62e+00 |
| 6 | 55 | 57.3 | 327 5 | Q93345 | C36B1.11 PROTEIN. | 3.62e+00 |
| 7 | 55 | 57.3 | 967 4 | Q94853 | KIA0753 PROTEIN. | 3.62e+00 |
| 8 | 54 | 56.3 | 108 10 | Q48607 | HYPOTHETICAL 12.3 KD P | 5.86e+00 |
| 9 | 54 | 56.3 | 238 11 | Q55075 | N-CADHERIN (FRAGMENT) | 5.86e+00 |
| 10 | 54 | 56.3 | 878 4 | Q15855 | UVOMORULIN PRECURSOR (| 5.86e+00 |
| 11 | 54 | 56.3 | 906 11 | Q921Y3 | TESTICULAR N-CADHERIN. | 5.86e+00 |
| 12 | 54 | 56.3 | 2258 14 | Q89238 | LARGE PROTEIN. | 5.86e+00 |
| 13 | 54 | 56.3 | 2258 14 | Q81081 | RNA POLYMERASE. | 5.86e+00 |
| 14 | 53 | 55.2 | 464 4 | Q43159 | KIA0409 (FRAGMENT). | 9.43e+00 |
| 15 | 53 | 55.2 | 775 10 | P93205 | SERINE PROTEASE, SBT2. | 9.43e+00 |
| 16 | 53 | 55.2 | 1980 14 | Q84185 | SENDAI VIRUS (STRAIN 2 | 9.43e+00 |
| 17 | 53 | 55.2 | 2223 14 | Q9WF25 | L PROTEIN. | 9.43e+00 |
| 18 | 53 | 55.2 | 2228 14 | Q98705 | RNA POLYMERASE PROTEIN | 9.43e+00 |
| 19 | 53 | 55.2 | 2228 14 | Q55530 | RNA POLYMERASE PROTEIN | 9.43e+00 |
| 20 | 53 | 55.2 | 2228 14 | Q55528 | RNA POLYMERASE PROTEIN | 9.43e+00 |

| RESULT | ID | PRELIMINARY: | PRT: | 993 AA. | RNA POLYMERASE I. | 9.43e+00 |
|--------|--|---|------|---------|------------------------|----------|
| AC | 035902 | | | | T10024.22 | 1.51e+01 |
| DT | 01-JAN-1998 | (TREMblrel. 05. Created) | | | HOLLIDAY JUNCTION DNA | 1.51e+01 |
| DT | 01-JAN-1998 | (TREMblrel. 05. Last sequence update) | | | ORF248. | 1.51e+01 |
| DT | 01-NOV-1998 | (TREMblrel. 12. Last annotation update) | | | F54D5.11 PROTEIN. | 1.51e+01 |
| DE | DESMOGLEIN 3 (FRAGMENT). | | | | N UTILIZATION SUBSTANC | 1.51e+01 |
| GN | DSG3. | | | | PROPENOXIDASE. | 1.51e+01 |
| OS | Mus musculus (Mouse). | | | | CALPAIN LP82. | 1.51e+01 |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; | | | | CALPAIN LP82. | 1.51e+01 |
| CC | Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | LENS-SPECIFIC CALPAIN | 1.51e+01 |
| CC | SEQUENCE FROM N.A. | | | | LENS-SPECIFIC CALPAIN | 1.51e+01 |
| CC | ISHIKAWA H., LI K., UIRTO J. | | | | LENS-SPECIFIC CALPAIN | 1.51e+01 |
| CC | Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases. | | | | CALPAIN 3 (EC 3.4.22.1 | 1.51e+01 |
| CC | -1 SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY). | | | | CALPAIN ISOPORN LP85. | 1.51e+01 |
| DR | EMBL: U86016; AAB65091.1; . | | | | MUSCLE TYPE CALPAIN P9 | 1.51e+01 |
| DR | HSSP: P15116; INCU. | | | | SKELETAL MUSCLE SPECIF | 1.51e+01 |
| DR | PROSITE: PS00232; CADHERIN; 2. | | | | CADHERIN-RELATED NEURA | 1.51e+01 |
| DR | PFAM: PF00028; cadherin; 4. | | | | HYPOTHETICAL PROTEIN (| 2.39e+01 |
| KW | Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat. | | | | COSMID T16A1. | 2.39e+01 |
| FT | NON_TER | | | | RIBOSOMAL PROTEIN S3. | 2.39e+01 |
| SO | SEQUENCE | | | | POLYMERASE MEMBRANE P | 2.39e+01 |

ALIGNMENTS

| RESULT | ID | PRELIMINARY: | PRT: | 993 AA. | RNA POLYMERASE I. | 9.43e+00 |
|--------|--|---|------|---------|------------------------|----------|
| AC | 035902 | | | | T10024.22 | 1.51e+01 |
| DT | 01-JAN-1998 | (TREMblrel. 05. Created) | | | HOLLIDAY JUNCTION DNA | 1.51e+01 |
| DT | 01-JAN-1998 | (TREMblrel. 05. Last sequence update) | | | ORF248. | 1.51e+01 |
| DT | 01-NOV-1998 | (TREMblrel. 12. Last annotation update) | | | F54D5.11 PROTEIN. | 1.51e+01 |
| DE | DESMOGLEIN 3 (FRAGMENT). | | | | N UTILIZATION SUBSTANC | 1.51e+01 |
| GN | DSG3. | | | | PROPENOXIDASE. | 1.51e+01 |
| OS | Mus musculus (Mouse). | | | | CALPAIN LP82. | 1.51e+01 |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; | | | | CALPAIN LP82. | 1.51e+01 |
| CC | Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | LENS-SPECIFIC CALPAIN | 1.51e+01 |
| CC | SEQUENCE FROM N.A. | | | | LENS-SPECIFIC CALPAIN | 1.51e+01 |
| CC | ISHIKAWA H., LI K., UIRTO J. | | | | LENS-SPECIFIC CALPAIN | 1.51e+01 |
| CC | Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases. | | | | CALPAIN 3 (EC 3.4.22.1 | 1.51e+01 |
| CC | -1 SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY). | | | | CALPAIN ISOPORN LP85. | 1.51e+01 |
| DR | EMBL: U86016; AAB65091.1; . | | | | MUSCLE TYPE CALPAIN P9 | 1.51e+01 |
| DR | HSSP: P15116; INCU. | | | | SKELETAL MUSCLE SPECIF | 1.51e+01 |
| DR | PROSITE: PS00232; CADHERIN; 2. | | | | CADHERIN-RELATED NEURA | 1.51e+01 |
| DR | PFAM: PF00028; cadherin; 4. | | | | HYPOTHETICAL PROTEIN (| 2.39e+01 |
| KW | Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat. | | | | COSMID T16A1. | 2.39e+01 |
| FT | NON_TER | | | | RIBOSOMAL PROTEIN S3. | 2.39e+01 |
| SO | SEQUENCE | | | | POLYMERASE MEMBRANE P | 2.39e+01 |

Query Match: 95.8%; Score 92; DB 11; Length 993;
Best Local Similarity 93.3%; Pred. No. 2.67e-09;
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

OC euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, COLUMBIA:
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUS S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.,
RT Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC004683; AAC28760.1;
SQ SEQUENCE 1421 AA; 154325 MW; 1B6D2FF1 CRC32;

Query Match
Best Local Similarity 63.5%; Score 61; DB 10; Length 1421;
Pred. No. 1.77e+01;
Matches 6; Conservative 8; Mismatches 1; Indels 0; Gaps 0;

Db 262 VDSKISYEITONPG 276
OY 1 LNSKIAFKIVSOEPA 15

RESULT 3
ID P72670 PRELIMINARY; PRT; 138 AA.
AC P72670;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 16.6 KD PROTEIN.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE: 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASANIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAKOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
DR EMBL: D90899; BAA16672.1;
KW Hypothetical protein.
SQ SEQUENCE 138 AA; 16580 MW; CA436352 CRC32;

Query Match
Best Local Similarity 60.4%; Score 58; DB 2; Length 138;
Pred. No. 8.20e-01;
Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 28 AKLAFEVIPODPT 40
OY 3 SKIAFKIVSOEPA 15

RESULT 4
ID Q20152 PRELIMINARY; PRT; 225 AA.
AC Q20152;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE F38B7.4 PROTEIN.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;

```

```

OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA LENNARD N.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAWTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HARKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIRKEN L., ROOPRA A., SAUNDERS D., SHOMKEEN R.,
RA SPALDON N., SMITH A., SONNHAMER E., STADEN R., SULSTON J.,
RA THERIERY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL: Z74033; CA98472.1;
SQ SEQUENCE 225 AA; 26356 MW; 66EPA455 CRC32;

Query Match
Best Local Similarity 57.3%; Score 55; DB 5; Length 225;
Pred. No. 3.62e+00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 205 LASKVSEFVRSQVPS 219
OY 1 LNSKIAFKIVSOEPA 15

RESULT 5
ID Q9YWS9 PRELIMINARY; PRT; 252 AA.
AC Q9YWS9;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ORF MSV032 HYPOTHETICAL PROTEIN.
GN MSV032.
OS Melanoplus sanguinipes entomopoxvirus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TUCSON;
RX MEDLINE: 99102612.
RA AFONSO C.L., TOLMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus."
RL J. Virol. 73:533-552(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-TUCSON;
RA AFONSO C.L., TOLMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF063866; AAC97615.1;
SQ SEQUENCE 252 AA; 29068 MW; 3A7E442E CRC32;

Query Match
Best Local Similarity 57.3%; Score 55; DB 14; Length 252;
Pred. No. 3.62e+00;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 202 SKVPFKILOES 213
OY 3 SKIAFKIVSOEP 14

RESULT 6
ID Q93345 PRELIMINARY; PRT; 327 AA.
AC Q93345;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

```

DT 01-JAN-1999 (TRENBLREL. 09, Last annotation update)
DE C36B1.11 PROTEIN.
GN C36B1.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditia; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA LENNARD N.
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COUSON A.,
RA CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.*
RL Nature 368:32-38(1994).
DR EMBL: 280215; CAB02276.1; -
SQ SEQUENCE 327 AA; 37139 MW; D8A2C259 CRC32;

Query Match 57.3%; Score 55; DB 5; Length 327;
Best Local Similarity 80.0%; Pred. No. 3.62e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 81 KIAFKIASOE 90
1:|||||111
4 KIAFKIASOE 13

RESULT 7
ID 094853 PRELIMINARY; PRT: 967 AA.
AC 094853:
DT 01-MAY-1999 (TRENBLREL. 10, Created)
DT 01-MAY-1999 (TRENBLREL. 10, Last sequence update)
DT 01-MAY-1999 (TRENBLREL. 10, Last annotation update)
DE KIA0753 PROTEIN.
GN KIA0753
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99087487.
RA NAGASE T., ISHIKAWA K., SUYAMA M., KIKUNO R., MIYAJIMA N., TANAKA A.,
RA KOTANI H., NOMURA N., OHARA O.,
RT *Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.*
RL DNA Res 5:277-286(1998).
DR EMBL: AB018295; BAA34473.1; -
SQ SEQUENCE 967 AA; 109376 MW; C8749C0A CRC32;

Query Match 57.3%; Score 55; DB 4; Length 967;
Best Local Similarity 46.2%; Pred. No. 3.62e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 83 LGSVSFSVISOE 95
1:|||||111
1 LNSKIAFKIVSOE 13

RESULT 8
ID 048607 PRELIMINARY; PRT: 108 AA.

AC 048607:
DT 01-JUN-1998 (TRENBLREL. 06, Created)
DT 01-JUN-1998 (TRENBLREL. 06, Last sequence update)
DT 01-AUG-1998 (TRENBLREL. 07, Last annotation update)
DE HYPOHETICAL 12.3 KD PROTEIN (FRAGMENT).
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Hordeum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ALBOSTRIANS, CV. HAISA; TISSUE-WHITE LEAF;
RA HESS W.R., GOLD R., BOERNER T.,
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ222776; CAI0981.1; -
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 108 AA; 12250 MW; E5FD23CE CRC32;

Query Match 56.3%; Score 54; DB 10; Length 108;
Best Local Similarity 45.3%; Pred. No. 5.86e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 15 LEFKINODPP 25
1:|||||111
5 IAFKIVSOEPA 15

RESULT 9
ID 05075 PRELIMINARY; PRT: 238 AA.
AC 05075:
DT 01-JUN-1998 (TRENBLREL. 06, Created)
DT 01-JUN-1998 (TRENBLREL. 06, Last sequence update)
DT 01-NOV-1999 (TRENBLREL. 12, Last annotation update)
DE N-CADHERIN (FRAGMENT).
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetus.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-Ovary;
RA LEVENBERG S., SADOT E., GOITCHEBERG P., GEIGER B.,
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AJ003143; CA05899.1; -
DR HSSP: P15116; INCD.
DR PROSITE: P50032; CADHERIN; 2.
KW PFAM: PF00028; cadherin; 2.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON_TER 1
FT NON_TER 238
SQ SEQUENCE 238 AA; 26234 MW; 0337A817 CRC32;

Query Match 56.3%; Score 54; DB 11; Length 238;
Best Local Similarity 40.0%; Pred. No. 5.86e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
Db 79 LNCMTYRILSOAPS 93
1:|||||111
1 LNSKIAFKIVSOEPA 15

RESULT 10
ID 015855 PRELIMINARY; PRT: 878 AA.
AC 015855: Q16194; Q13799;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-NOV-1999 (TRENBLREL. 12, Last annotation update)
DE UVOMORULIN PRECURSOR (E-CADHERIN) (ARC-1/UVOMORULIN).
GN UV0.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.

RN [1]
 SEQUENCE FROM N.A.
 RC TISSUE-LIVER;
 RX MEDLINE: 94242050.
 RA RIMM D.L., MORROW J.S.;
 RT "Molecular cloning of human E-cadherin suggests a novel subdivision of
 the cadherin superfamily."
 RL Biochem. Biophys. Res. Commun. 200:1754-1761(1994).
 RN [2]
 SEQUENCE OF 333-472 FROM N.A.
 RX MEDLINE: 94306394.
 RA BECKER K.F., ATKINSON M.J., REICH U., BECKER I., NEKARDA H.,
 RA SIEMER J.R., HOFER H.;
 RT "E-cadherin gene mutations provide clues to diffuse type gastric
 carcinomas."
 RL Cancer Res. 54:3845-3852(1994).
 RN [3]
 SEQUENCE OF 261-388 FROM N.A.
 RC TISSUE-LIVER;
 RA FRIKEN U.H.;
 RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: S72492; AAD14108.1; JOINED.
 DR EMBL: S72491; AAD14108.1; JOINED.
 DR EMBL: L08599; AAA61259.1; JOINED.
 DR EMBL: X52379; CA36522.1; JOINED.
 DR HSER; P09603; ISUH.
 DR PROSITE: P500232; CADHERIN; 3.
 DR PFAM: PF00028; cadherin; 5.
 DR PFAM: PF01049; Cadherin_C-term; 1.
 DR PRINTS: PR00205; CADHERIN.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 878 UVOMORULIN.
 FT SO SEQUENCE 878 AA; 96741 MW; 8E7F0180 CRC32;
 Query Match 56.3%; Score 54; DB 4; Length 878;
 Best Local Similarity 53.8%; Pred. No. 5.86e+00;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 DB 293 NAAIATYILSDP 305
 QY 2 LNSKIAFKIVSOEPA 14
 RESULT 11
 ID 0921Y3 PRELIMINARY; PRT; 906 AA.
 AC 0921Y3
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE TESTICULAR N-CADHERIN.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY;
 RX MEDLINE: 98187820.
 RA CHUNG S.S., MO M.Y., SILVESTRINI B., LEE W.M., CHENG C.Y.;
 RT "Rat testicular N-cadherin: its complementary deoxyribonucleic acid
 cloning and regulation."
 RL Endocrinology 139:1853-1862(1998).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF097593; AAC83818.1; JOINED.
 DR HSP: P15116; INCI
 DR PROSITE: P500232; CADHERIN; 3.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SO SEQUENCE 906 AA; 99686 MW; 6C52E299 CRC32;
 Query Match 56.3%; Score 54; DB 11; Length 906;
 Best Local Similarity 40.0%; Pred. No. 5.86e+00;

Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
 DB 300 LNSKIAFKIVSOEPA 314
 QY 1 LNSKIAFKIVSOEPA 15
 RESULT 12
 ID 089238 PRELIMINARY; PRT; 2258 AA.
 AC 089238;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE LARGE PROTEIN.
 OS Human parainfluenza virus 3.
 CC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 CC Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN-GP;
 RX MEDLINE: 98436895.
 RA OHSAWA K., YAMADA A., TAKEUCHI K., WATANABE Y., MIYATA H., SATO H.;
 RT "Genetic characterization of parainfluenza virus 3 derived from guinea
 pig."
 RL J. Vet. Med. Sci. 60:919-922(1998).
 DR EMBL: AB012132; BA32575.1; JOINED.
 DR PFAM: PF00946; Paramyx_RNA_pol; 1.
 SO SEQUENCE 2258 AA; 258800 MW; 8E00950F CRC32;
 Query Match 56.3%; Score 54; DB 14; Length 2258;
 Best Local Similarity 33.3%; Pred. No. 5.86e+00;
 Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;
 DB 998 LDRSVLYRIMNOEPG 1012
 QY 1 LNSKIAFKIVSOEPA 15
 RESULT 13
 ID 081081 PRELIMINARY; PRT; 2258 AA.
 AC 081081;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE RNA POLYMERASE.
 GN L.
 OS Human parainfluenza virus 3.
 CC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 CC Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN-JS;
 RX MEDLINE: 94091054.
 RA STOKES A., TIERNEY E.L., SARRIS C.M., MURPHY B.R., HALL S.L.;
 RT "The complete nucleotide sequence of two cold-adapted, temperature-
 sensitive attenuated mutant vaccine viruses (cp12 and cp45) derived
 from the JS strain of human parainfluenza virus type 3 (PIV3)."
 RL Virus Res. 30:43-52(1993).
 RN [2]
 SEQUENCE FROM N.A.
 RC STRAIN-JS;
 RA DUBBIN A.D.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 SEQUENCE FROM N.A.
 RC STRAIN-JS;
 RA DUBBIN A.D.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U51116; ABA4690.1; JOINED.
 DR PFAM: PF00946; Paramyx_RNA_pol; 1.
 SO SEQUENCE 2258 AA; 258927 MW; 4B7A664B CRC32;
 Query Match 56.3%; Score 54; DB 14; Length 2258;

Best Local Similarity 33.3%: Pred. No. 5.66e+00;
Matches 5: Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 998 LDRSVLYRINNOEPC 1012

OY 1 LNSKIAFKIVSOEPA 15

Db 734 LSKVTEFKTVSRÖKA 748
OY 1 LNSKIAFKIVSOEPA 15

Search completed: Sat May 13 07:10:03 2000
Job time : 152 secs.

RESULT 14

ID 043159 PRELIMINARY: PRT: 464 AA.

AC 043159:

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)

DE KIAA0409 (FRAGMENT).

GN KIAA0409.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Homiidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN: NAGASE T., NAKAJIMA D., SEKI N., OHIRA M., MIYAJIMA N.,

RA ISHIIKAWA K., KOTANI H., NOMURA N., OHARA O.;

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB007869; BAA23705.1; -

FT NON_TER

SQ SEQUENCE 464 AA: 51479 MW: 4CDDDBCE CRC32:

Query Match

Best Local Similarity 55.2%: Score 53; DB 4; Length 464;

Matches 6: Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 416 TKLGFKIVSKD 426

OY 3 SKIAFKIVSOE 13

RESULT 15

ID P93205 PRELIMINARY: PRT: 775 AA.

AC P93205:

DT 01-MAY-1997 (TREMBlrel. 03, Created)

DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)

DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)

DE SERINE PROTEASE, SBT2.

GN SBT2.

OS Lycopersicon esculentum (Tomato).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;

OC core eudicots; Asterales; euasterids I; Solanales; Solanaceae;

OC Solanum.

RN [1]

RP SEQUENCE FROM N.A.

RA SCHALLER A.;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. VFWS;

RA MEICHTREY J., AMRHEIN N., SCHALLER A.;

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: X98930; CAA67430.1; -

DR EMBL: AJ006379; CAA07000.1; -

DR HSSP: Q99405; 1MPT.

DR MENDEL: 8813; Lyces:1086;8813.

DR PFAM: PF00082; Peptidase_S8: 3.

DR PRINTS: PR00723; SUBTILISIN.

KW Protease; Serine protease.

SQ SEQUENCE 775 AA: 83115 MW: 2EBF3C5E CRC32:

Query Match

Best Local Similarity 55.2%: Score 53; DB 10; Length 775;

Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

THIS PAGE BLANK (USPTO)

M P E R E H
(TM)

Release 3.1A John F. Collins, Bioinformatics Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:16:39 2000: Maspar time 3.02 Seconds
Tabular output not generated. 117,602 Million cell updates/sec

Title: >US-08-991-628-4
Description: (1-15) from US08991628.pep
Perfect Score: 108
Sequence: 1 TPFMLSRNTGEVRT 15

Scoring table: PAM 150
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a:geneseq35
1:geneseqp

Statistics: Mean 19.380; Variance 58.525; scale 0.331

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|--------|------------------------|-----------|
| 1 | 108 | 100.0 | 15 | W04844 | Self epitope of desmog | 6.82e-05 |
| 2 | 108 | 100.0 | 15 | W64816 | Desmoglein 3 206-220 | 6.82e-05 |
| 3 | 108 | 100.0 | 15 | W78815 | Desmoglein 3 protein f | 6.82e-05 |
| 4 | 108 | 100.0 | 614 | W07908 | Pemphigus vulgaris ant | 6.82e-05 |
| 5 | 108 | 100.0 | 999 | R30742 | Human pemphigus vulgar | 6.82e-05 |
| 6 | 81.5 | 778 | 1 | W15489 | Pemphigus foliaceus an | 1.76e-02 |
| 7 | 61 | 56.5 | 583 | R07999 | Asparagine synthetase | 2.03e-01 |
| 8 | 61 | 56.5 | 583 | R07998 | Asparagine synthetase | 2.03e-01 |
| 9 | 60 | 55.6 | 163 | W38654 | S. pneumoniae aspartat | 2.03e-01 |
| 10 | 59 | 54.6 | 616 | R68667 | Rat pneumococcal p55 | 3.31e-01 |
| 11 | 57 | 52.8 | 263 | W13010 | Segment of desmosomal | 5.39e-01 |
| 12 | 57 | 52.8 | 560 | W13009 | Segment of desmosomal | 5.39e-01 |
| 13 | 55 | 50.9 | 660 | R69633 | Human interleukin-12 r | 8.73e-01 |
| 14 | 55 | 50.9 | 662 | R69632 | Human interleukin-12 r | 8.73e-01 |
| 15 | 55 | 50.9 | 662 | W12772 | Human interleukin-12 b | 8.73e-01 |
| 16 | 54 | 50.0 | 325 | W92443 | Mouse MC5 protein. | 1.11e-02 |
| 17 | 54 | 50.0 | 325 | W87870 | Mouse melanocortin rec | 1.11e-02 |
| 18 | 54 | 50.0 | 325 | W79688 | Melanocortin-5 recepto | 1.11e-02 |
| 19 | 54 | 50.0 | 325 | W19705 | Melanocortin-5 recepto | 1.11e-02 |
| 20 | 54 | 50.0 | 325 | W33735 | Mouse melanocortin-5 r | 1.11e-02 |
| 21 | 54 | 50.0 | 325 | R79501 | Rat melanocortin recep | 1.11e-02 |
| 22 | 54 | 50.0 | 325 | W41067 | Mouse melanocortin-5 r | 1.11e-02 |
| 23 | 54 | 50.0 | 325 | W37853 | Mouse melanocortin-5 r | 1.11e-02 |

| | | | | | | | |
|----|----|------|------|---|--------|-------------------------|----------|
| 24 | 54 | 50.0 | 433 | 1 | W00365 | Human cyclin B1. | 1.11e-02 |
| 25 | 54 | 50.0 | 555 | 1 | W85133 | A desaturase enzyme en | 1.11e-02 |
| 26 | 54 | 50.0 | 555 | 1 | W84154 | Human desaturase enzym | 1.11e-02 |
| 27 | 54 | 50.0 | 608 | 1 | W85134 | A desaturase enzyme en | 1.11e-02 |
| 28 | 54 | 50.0 | 608 | 1 | W84155 | Human desaturase enzym | 1.11e-02 |
| 29 | 54 | 50.0 | 746 | 1 | W85135 | A desaturase enzyme en | 1.11e-02 |
| 30 | 54 | 50.0 | 746 | 1 | W84156 | Human desaturase enzym | 1.11e-02 |
| 31 | 53 | 49.1 | 609 | 1 | P30206 | Sequence encoded by pf | 1.40e-02 |
| 32 | 51 | 47.2 | 400 | 1 | R10919 | Human GM-CSF receptor. | 2.24e-02 |
| 33 | 51 | 47.2 | 552 | 1 | R24017 | Fusion protein GM-CSF r | 2.24e-02 |
| 34 | 51 | 47.2 | 1686 | 1 | W70991 | Human class II p13 kin | 2.24e-02 |
| 35 | 51 | 47.2 | 1726 | 1 | W38756 | Phosphatidyl inositol | 2.24e-02 |
| 36 | 50 | 46.3 | 166 | 1 | P70358 | Sequence encoded by pl | 2.83e-02 |
| 37 | 50 | 46.3 | 193 | 1 | P60877 | Hook region #8 contig 1 | 2.83e-02 |
| 38 | 50 | 46.3 | 225 | 1 | P70061 | Sequence of human grow | 2.83e-02 |
| 39 | 50 | 46.3 | 297 | 1 | W87867 | Bovine melanocortin re | 2.83e-02 |
| 40 | 50 | 46.3 | 297 | 1 | W92440 | Human MC2 protein. | 2.83e-02 |
| 41 | 50 | 46.3 | 472 | 1 | W40072 | Human retinoid recepto | 2.83e-02 |
| 42 | 50 | 46.3 | 720 | 1 | W19266 | Lactobacillus amylovor | 2.83e-02 |
| 43 | 50 | 46.3 | 878 | 1 | R85487 | Human E-cadherin precu | 2.83e-02 |
| 44 | 50 | 46.3 | 916 | 1 | Y10967 | H. pylori ORF hppe1521 | 2.83e-02 |
| 45 | 50 | 46.3 | 928 | 1 | W21017 | H. pylori cytoplasmic | 2.83e-02 |

ALIGNMENTS

RESULT 1
ID W04844; standard; peptide: 15 AA.
AC W04844;
DE 18-FEB-1997 (first entry)
KW Self epitope of desmoglein 3, implicated in autoimmune disease.
KW Tolerisation, self-epitope; antigen; autoimmune disease.
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW pemphigus vulgaris; desmoglein; multiple sclerosis;
KW herpes simplex virus; adenovirus; phosphonomomutase;
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW Influenza; haemagglutinin; reovirus; sigma protein.
OS Homo sapiens.
PN W09627387-AL.
PD 12-SEP-1996.
PF 07-MAR-1996; U03182.
PR 07-MAR-1995; US-400796.
PI (HARD) HARVARD COLLEGE.
PI Strominger JL, Nucleic Acids Res.
DR WPI; 96-425218/42.
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT antigens - useful in disease treatment, and method for
PT identification of other self and non-self antigens implicated in
PT auto-immune disease
PS Claim 1: Page 40; 58pp; English.
CC Pharmaceutical preparations for tolerisation to antigens comprise
CC either an isolated human non-collagen or non-mysin basic protein
CC (MBP) polypeptide which is capable of tolerising an individual to an
CC autoantigen; or an isolated human pathogen polypeptide capable of
CC tolerising an individual to that polypeptide. In both cases, the
CC polypeptide (whether self or non-self) includes an amino acid
CC sequence corresponding to a sequence motif for a MHC class II
CC protein, such as HLA-DR, which is associated with a human autoimmune
CC disease and which binds to the polypeptide to activate autoreactive
CC T-cells in individuals with the autoimmune disease. This peptide is
CC derived from the human desmoglein 3 protein (amino acids 206-220)
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC derived from the human desmoglein protein are described in W04844-47.
SQ Sequence 15 AA;

Query Match 100.0%; Score 108; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.82e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 TPFMLSRNTGEVRT 15
OY 1 TPFMLSRNTGEVRT 15

RESULT 2
ID W64816 standard; peptide: 15 AA.
AC W64816:
DT 29-SEP-1998 (first entry)
DE Desmoglein-3 206-220.
KW Desmoglein; DG: gene therapy; pemphigus vulgaris; microparticle;
KW autoantigen; autoimmune disease; MHC.
OS Homo sapiens.
PN US5783567-A.
PD 21-JUL-1998.
PF 22-JAN-1997: 787547.
PR 22-JAN-1997: US-787547.
PA (PANG-) PANGAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS:
DR WPI: 98-427077/36.
PT Microparticle encapsulated nucleic acids - for recombinant
PS expression of proteins e.g. in gene therapy.
CC Disclosure: Column 4: 42pp; English.
CC The patent describes a new preparation of microparticles each
CC comprising a polymeric matrix and a nucleic acid. The polymeric
CC matrix consists of one or more synthetic polymers having a solubility
CC in water of less than 1 mg/l (e.g. poly-lactic-co-glycolic acid);
CC and at least 90% of the microparticles have a diameter of less than
CC 100 microns. The microparticles are useful for the delivery of nucleic
CC acids to phagocytic cells. In one embodiment the microparticles are
CC less than 20 microns in diameter and the nucleic acid (preferably in
CC closed circular form) includes an expression control sequence
CC operatively linked to a coding sequence, where the expression product
CC of the coding sequence is a polypeptide having a length and a sequence
CC which permits it to bind to an MHC class I or II molecule. The
CC expression product is thus an effective stimulator of an immune
CC response in mammals. The present sequence, an antigenic portion of
CC desmoglein 3, is an example of an MHC class II peptide which can be
CC expressed by the nucleic acid. It is associated with pemphigus
CC vulgaris.
SQ Sequence 15 AA:

Query Match 100.0%; Score 108; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.82e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TPFLLSRNTGEVRT 15
Qy 1 TPFLLSRNTGEVRT 15

RESULT 3
ID W78815 standard; peptide: 15 AA.
AC W78815:
DT 17-NOV-1998 (first entry)
DE Desmoglein 3 protein fragment 206-220.
KW Microparticle; delivery; polymeric matrix; autoantigen; tumour antigen;
KW class II associated peptide; pathogen; gene therapy; genetic disease;
KW infection; downregulation; immune response.
OS Homo sapiens.
PN WO9831598-A1.
PD 23-JUL-1998.
PF 22-JAN-1998: U01499.
PR 06-JAN-1998: US-003253.
PR 22-JAN-1997: US-787547.
PA (PANG-) PANGAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS, Lunsford LB:
DR WPI: 98-427556/36.
PT New preparations of microparticles - comprising a synthetic polymer
PT matrix and nucleic acid comprising an expression vector for use in
PT gene therapy.
PS Disclosure: Page 8: 101pp; English.
CC A microparticle preparation (MP) has been developed, consisting of
CC microparticles having a diameter of less than 100 mu m. The MP comprises:
CC (a) a polymeric matrix (PM) consisting of one or more synthetic polymers
CC having a solubility in water of less than 1 mg/l; and (b) an expression

CC vector selected from RNA molecules (at least 50% of which are closed
CC circles) or circular plasmid DNA (at least 50% of which are supercoiled).
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)
CC a PM; and (b) a NAM comprising an expression control sequence operatively
CC linked to a coding sequence, where the coding sequence encodes an
CC expression product selected from: (1) a polypeptide at least 7 amino
CC acids in length, having a sequence identical to the sequence of: (1) a
CC fragment of a naturally-occurring mammalian protein; or (11) a fragment
CC of a naturally-occurring protein from an infectious agent which infects
CC a mammal; (2) a peptide having a length and sequence which permits it to
CC bind to an MHC class I or II molecule; and (3) the polypeptide or the
CC peptide linked to a trafficking sequence. W69763 to W69765, and W78793
CC to W78897 are peptide fragments for use in the present invention. The
CC MPs are highly effective vehicles for the delivery of polynucleotides
CC into phagocytic cells. They can be used for gene therapy, e.g. for
CC treating genetic diseases, infections or tumours or for downregulating
CC an immune response.
SQ Sequence 15 AA:

Query Match 100.0%; Score 108; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.82e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TPFLLSRNTGEVRT 15
Qy 1 TPFLLSRNTGEVRT 15

RESULT 4
ID W07908 standard; protein: 614 AA.
AC W07908:
DT 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; Immunoglobulin G; IgG1; fusion protein; diagnosis;
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.
OS Homo sapiens.
PN J08188540-A.
PD 23-JUL-1996.
PF 30-JUN-1995: 165632.
PR 30-JUN-1994: JP-173291.
PA (NISHU) NISHUKAWA T.
DR WPI: 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris auto-antibody -
PT useful to treat and diagnose pemphigus vulgaris
PS Claim 1; Page 7-9; 9pp; Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose
CC pemphigus vulgaris.
SQ Sequence 614 AA:

Query Match 100.0%; Score 108; DB 1; Length 614;
Best Local Similarity 100.0%; Pred. No. 6.82e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 205 TPFLLSRNTGEVRT 219
Qy 1 TPFLLSRNTGEVRT 15

RESULT 5
ID R30742 standard; protein: 999 AA.
AC R30742:
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130KD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.

PN US798918-A.
 PD 15-DEC-1992.
 PF 27-NOV-1991: 798918.
 PR 27-NOV-1991: US-798918.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PI Amagai M, Klaus-Kovtun V, Stanley JR;
 DR WPI: 93-067436/08.
 DR N-PSDB: 035992.
 PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
 PT diagnostic and therapeutic uses
 PS Disclosure; Fig 7; 50pp; English.
 CC This sequence is the pemphigus vulgaris 130KD antigen. The protein
 CC and its encoding DNA may be used in the diagnosis and treatment of
 CC pemphigus vulgaris. It is thought that the antigen may be a cell
 CC adhesion molecule.
 SO Sequence 999 AA;

Query Match 100.0%; Score 108; DB 1; Length 999;
 Best Local Similarity 100.0%; Pred. No. 6.82e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 206 TPMFLSRNTEGVRT 220
 |||||
 1 TPMFLSRNTEGVRT 15

RESULT 6
 ID W15489 standard; Protein: 778 AA.
 AC W15489;
 DT 17-JUN-1997 (first entry)
 DE Pemphigus foliaceus antigen-IgG constant region fusion protein.
 KM Pemphigus foliaceus: autoantibody; constant region; IgG;
 KW extracellular region; antigen; hinge portion; skin;
 OS dermatitis herpetiformis; fusion protein; detection; ss.
 FH Chimeric - Homo sapiens.
 FT Key Location/Qualifiers
 FT domain 1..545
 PN J09077800-A.
 PD 25-MAR-1997.
 PF 12-SEP-1995: 260899
 PR 12-SEP-1995: JP-260899.
 PA (NISH/) NISHIKAWA T.
 DR WPI: 97-241758/22.
 DR P-PSDB: T66428.
 PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
 PT through the hinge region used to treat pemphigus foliaceus
 PS Claim 1: Page 10-12; 17pp; Japanese.
 CC This sequence represents a fused protein recognised by pemphigus
 CC foliaceus patient autoantibody which comprises the constant region
 CC of IgG linked to the extracellular region of pemphigus foliaceus
 CC antigen protein through the hinge portion. Pemphigus foliaceus is
 CC a chronic, generalised, vesicular and scaling skin eruption similar
 CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion
 CC protein is useful to treat pemphigus foliaceus. The antigen is
 CC especially administered through an adsorbent upon which the fusion
 CC protein is immobilised via a carrier. The fusion protein is also
 CC useful for detecting pemphigus foliaceus antibodies which is useful
 CC in immunodiagnosis. The fusion protein has little or no side effects.
 SO Sequence 778 AA;

Query Match 81.5%; Score 88; DB 1; Length 778;
 Best Local Similarity 66.7%; Pred. No. 1.76e-02;
 Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 206 SPMFIINRNGEVRT 220
 :|||||
 1 TPMFLSRNTEGVRT 15

RESULT 7
 ID R07999 standard; Protein: 583 AA.
 AC R07999;

DT 26-FEB-1991 (first entry)
 DE Asparagine synthetase AS2
 KM asparagine synthetase; transgenic plant; herbicide resistance;
 KW drought tolerance; nitrogen fixation; pea.
 OS Pisum sativum.
 FH Key Location/Qualifiers
 FT binding-site 1..4
 FT /label= glutamine binding site
 FT /note= "conserved in human AS and pea AS1 and 2."
 PN W09013633-A.
 PD 15-NOV-1990.
 PF 02-MAY-1990: U02443.
 PR 03-MAY-1989: US-347302.
 PR 26-APR-1990: US-514816.
 PA (YURO-) ROCKFELLER UNIV.
 PI Coruzzi GM, Tsai FY;
 DR WPI: 90-361471/48.
 DR N-PSDB: 006599.
 PT Plant asparagine synthetase - includes DNA expression system and
 PT transgenic plants
 PS Disclosure; Fig 2B; 91pp; English.
 CC The DNA sequence encoding this protein was isolated from cDNA
 CC clones selected from a pea nodule cDNA library from the "Sparkle"
 CC variety of P.sativum. Pea AS1 cDNA was used as a probe.
 CC The protein is produced by expression vectors containing the AS2
 CC coding sequence. Recombinant AS can be used to engineer herbicide
 CC resistance, as a dominant selectable marker, to select novel
 CC herbicides or compounds useful for synchronising plant cells in
 CC culture, etc.
 CC A comparison of pea AS and human AS polypeptides reveals an overall
 CC homology of 47% at the amino acid level. C.f. 86% between AS1 and
 CC AS2. There are several regions of high local homology (greater than
 CC 80%) shared between the pea AS and human AS polypeptides.
 SO See also Q06598, Q06622 and Q06623.
 SQ Sequence 583 AA;

Query Match 56.5%; Score 61; DB 1; Length 583;
 Best Local Similarity 77.8%; Pred. No. 2.03e+01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPMFLMSRK 330
 |||||
 1 TPMFLSRN 9

RESULT 8
 ID R07998 standard; Protein: 586 AA.
 AC R07998;
 DT 26-FEB-1991 (first entry)
 DE Asparagine synthetase AS1.
 KM asparagine synthetase; transgenic plant; herbicide resistance;
 KW drought tolerance; nitrogen fixation; pea.
 OS Pisum sativum.
 FH Key Location/Qualifiers
 FT region 116..128
 FT /label= 80% homology with human AS1 sequence
 FT region 218..243
 FT /label= 80% homology with human AS1
 FT region 340..348
 FT /label= 80% homology with human AS1
 FT region 352..360
 FT /label= 80% homology with human AS1
 FT region 392..401
 FT /label= 80% homology with human AS1
 FT region 486..500
 FT /label= 80% homology with human AS1
 FT binding-site 1..4
 FT /label= putative glutamine binding site
 PN W09013633-A.
 PD 15-NOV-1990.
 PF 02-MAY-1990: U02443.
 PR 03-MAY-1989: US-347302.
 PR 26-APR-1990: US-514816.

PA (UVR0-) ROCKFELLER UNIV.
 PI CORUZ21 GM, Tsal, FY;
 DR WPI: 90-361471/48.
 DR N-PSDB: 006599.
 PT Plant asparagine synthetase - includes DNA expression system and
 PS transgenic plants
 PS Disclosure: fig 2A: 91pp: English.
 CC The DNA sequence encoding this protein was isolated from cDNA
 CC clones selected from a pea node cDNA library from the "Sparkle"
 CC variety of *P. sativum*. Human AS cDNA was used as a probe.
 CC The protein is produced by expression vectors containing the AS1
 CC coding sequence. Recombinant AS can be used to engineer herbicide
 CC resistance, as a dominant selectable marker, to select novel
 CC herbicides or compounds useful for synchronising plant cells in
 CC culture, etc.
 CC See also 006599, 006622 and 006623.
 SO Sequence 586 AA;

Query Match 56.5%; Score 61; DB 1; Length 586;
 Best Local Similarity 77.8%; Pred. No. 2,03e+01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 323 TPFLLMSRK 331
 |||||:
 OY 1 TPFLLSRN 9

RESULT 9
 ID W38654 standard; Protein: 163 AA.
 AC W38654;
 DT 09-NOV-1998 (first entry)
 DE S. pneumoniae aspartate ammonia lyase.
 KM Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 KM immunological response; inoculation; antibody production; inhibitor;
 KM T cell immune response; antimicrobial compound; bacterial adhesion;
 KM extracellular matrix protein; protein-mediated cell invasion; wound;
 KM pathogenesis.
 OS Streptococcus pneumoniae.
 PN MO9743303-A1.
 PD 20-NOV-1997.
 PR 14-MAY-1997: U07950.
 PR 14-MAY-1996: US-017670.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
 PI Stodola RK;
 DR WPI: 98-008793/01.
 DR N-PSDB: T98699.
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
 PT diagnosing anti-microbial agents for treatment of bacterial
 PT infections
 PS Claim 12: Page 407: 483pp: English.
 CC This sequence represents a Streptococcus pneumoniae protein that, based
 CC on homology with an Escherichia coli protein, is a aspartate ammonia
 CC lyase, and is encoded by a DNA sequence of the invention.
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
 CC invention can be used to identify compounds which interact with and
 CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of
 CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.
 SO Sequence 163 AA;

Query Match 55.6%; Score 60; DB 1; Length 163;
 Best Local Similarity 60.0%; Pred. No. 2.59e+01;
 Matches 9; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

DB 115 AMFLLRKRGCEVQT 129
 :||| |:
 OY 2 PMFLLS-RNTGEVRT 15

RESULT 10
 ID R86867 standard; Protein: 616 AA.
 AC R86867;
 DT 27-AUG-1996 (first entry)
 DE Rat protocadherin pc5.
 KM Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
 KM catenin; therapy.
 OS Rattus rattus.
 PN MO9600289-A1.
 PD 04-JAN-1996.
 PR 26-JUN-1995: U08071.
 PR 27-JUN-1994: US-268161.
 PA (DOHE-) DOHERTY EYE INST.
 PI Suzuki S;
 DR WPI: 96-068873/07.
 DR N-PSDB: T03574.
 PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
 PT pc5 - involved in cell-cell adhesion and regulation activities
 PS Claim 17: Page 130-132: 146pp: English.
 CC R86865-R86867 represent the sequences for three protocadherins. This
 CC sequence represents the rat protocadherin pc5. These sequences are
 CC related to cadherin, and possess cell adhesive ability. Cadherins are
 CC glycosylated integral membrane proteins that are involved in cell-cell
 CC adhesion. Cadherins are composed of an N-terminal extracellular domain
 CC which consists of 5 unique subdomains, a membrane spanning domain, and a
 CC C-terminal cytoplasmic domain. The cytoplasmic domain interacts with the
 CC cytoskeleton through catenins and other cytoskeleton associated proteins.
 CC The cytoplasmic domain is not present in all cadherins, but in those
 CC which possess it, it is essential for the cadherin adhesive function.
 CC The cadherins which do not possess a cytoplasmic domain appear to
 CC function via a different method from those with a cytoplasmic domain.
 CC These sequences were isolated using primers 1 and 2 (see T03575 and
 CC T03576) The proteins may have regulatory functions in the cell, as well
 CC as the cell-cell adhesive properties. Antibodies produced against these
 CC sequences are useful for modulating the binding activity of these
 CC protocadherins, and can be used therapeutically.
 SO Sequence 616 AA;

Query Match 54.6%; Score 59; DB 1; Length 616;
 Best Local Similarity 61.5%; Pred. No. 3.31e+01;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 85 TPLFLLNTEGEI 97
 |||||:
 OY 1 TPFLLSRNTEV 13

RESULT 11
 ID W13010 standard; Protein: 263 AA.
 AC W13010;
 DT 21-NOV-1997 (first entry)
 DE Segment of desmosomal cadherin, desmoglein Dsg2.
 KM Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
 KM carcinoma; desmosome; antibody; epitope; diagnosis; detection;
 KM microemulsions; separation; enrichment; targeted delivery;
 KW metastatic.
 OS Homo sapiens.
 PN DE19531033-A1.
 PD 27-FEB-1997.
 PR 23-AUG-1995: 031033.
 PR 23-AUG-1995: DE-031033.
 PA (PROG-) PROGEM BIOTECHNIK GMBH.
 PI Franke WW, Schaefer S;
 DR WPI: 97-146518/14.

PT Antibody reactive with part of desmosomal cadherin - exposed on surface of epithelial or carcinoma cells, not bound to desmosomes, useful for diagnosis and treatment of carcinoma micrometastases

PS Claim 9: Page 5: 8pp: German.

CC The present sequence is a segment of the desmosomal cadherin (DC), desmoglein Dsg2, which is exposed on the surface of epithelial or carcinoma cells and not bound to desmosomes. An antibody (Ab) directed against epitopes of the present sequence can be used to diagnose, i.e. to detect carcinoma cells, especially micrometastases, not bound to desmosomes, to separate, enrich or detect living or fixed carcinoma cells by cell sorting methods and as a therapeutic to deliver agents, e.g. other Ab or toxins, to target cells. The Ab provides rapid and reliable detection of metastatic carcinoma, and detects parts of DC that are not accessible in desmosome bound cells, as in normal tissue or carcinomas.

CC Sequence 263 AA:

Query Match 52.8%; Score 57; DB 1; Length 263;
Best Local Similarity 50.0%; Pred. No. 5.39e+01;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 58 PVFYLKDTGET 69
OY 2 PMFLSRNTGEV 13

RESULT 12
ID W13009 standard; protein: 560 AA.

AC W13009:
DT 21-NOV-1997 (first entry)
DE Segment of desmosomal cadherin, desmoglein Dsg2.
KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial; carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW micrometastasis; separation; enrichment; targeted delivery;
KW metastatic.
OS Homo sapiens.
PN DE19531033-A1.
PD 27-FEB-1997.
PE 23-AUG-1995; 031033.
PR 23-AUG-1995; DE-031033.
PA (PROG-) PROGEN BIOTECHNIK GMBH.
PI Franke WW, Schaefer S;
DR WPI: 97-146518/14.
PT Antibody reactive with part of desmosomal cadherin - exposed on surface of epithelial or carcinoma cells, not bound to desmosomes, useful for diagnosis and treatment of carcinoma micrometastases

PS Claim 7: Page 5: 8pp: German.

CC The present sequence is a segment of the desmosomal cadherin (DC), desmoglein Dsg2, which is exposed on the surface of epithelial or carcinoma cells and not bound to desmosomes. An antibody (Ab) directed against epitopes of the present sequence can be used to diagnose, i.e. to detect carcinoma cells, especially micrometastases, not bound to desmosomes, to separate, enrich or detect living or fixed carcinoma cells by cell sorting methods and as a therapeutic to deliver agents, e.g. other Ab or toxins, to target cells. The Ab provides rapid and reliable detection of metastatic carcinoma, and detects parts of DC that are not accessible in desmosome bound cells, as in normal tissue or carcinomas.

CC Sequence 560 AA:

Query Match 52.8%; Score 57; DB 1; Length 560;
Best Local Similarity 50.0%; Pred. No. 5.39e+01;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 160 PVFYLKDTGET 171
OY 2 PMFLSRNTGEV 13

RESULT 13
ID R69633 standard; protein: 660 AA.

AC R69633:
DT 05-SEP-1995 (first entry)
DE Human interleukin-12 receptor.
KW Interleukin-12 receptor; IL-12; immune suppression;
KW immunosuppressive; graft-versus-host reaction; allograft rejection;
KW inflammation; autoimmune disease.
OS Homo sapiens.

PM Key Location/Qualifiers
FT peptide 1..23
FT /label- Sig_peptide 24..540
FT /note- "extracellular region"

FT region 52..64
FT /note- "sequence motif of cytokine receptor superfamily Cys52..Cys62Sw"

FT modified_site 121..123
FT /label- N-linked-glycosylation 222..226
FT region
FT /note- "cytokine receptor superfamily motif (W225KWS)"

FT modified_site 329..331
FT /label- N-linked-glycosylation 346..348
FT modified_site 352..354
FT /label- N-linked-glycosylation 442..444
FT modified_site 456..458
FT /label- N-linked-glycosylation 541..570
FT region
FT /label- Transmembrane_region 571..562
FT region
FT /label- Cytoplasmic_tail 577..584
FT region
FT /label- Cytoplasmic_tail 618..629
FT /note- "conserved area of cytoplasmic tail"

FT region
FT /label- Cytoplasmic_tail
FT /note- "conserved area of cytoplasmic tail"

FT EP-638644-A.
PD 15-FEB-1995.
PE 08-JUL-1994; 110657.
PR 19-JUL-1993; US-094649.
PR 31-MAY-1994; US-094713.
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
PI Chizzonite RA, Chua AO, Gubler UA, Trullt TP;
DR WPI: 95-076349/11.
PT DNA encoding a low affinity interleukin-12 receptor - used to bind or scavenge IL-12 to cause immune suppression, e.g. to suppress graft-vs-host reaction, allograft rejection or inflammation, and to treat autoimmune conditions

PS Claim 4: Page 31-33; 61pp: English.

CC A cDNA library of PMA-activated peripheral blood mononuclear cells in vector per-BOS was screened for interleukin-12 (IL-12) receptor cDNAs by panning. One isolated cDNA (683844) encoded a 662-amino acid low affinity IL-12 receptor (R69632). Another cDNA contained an additional 202 bp of 3' UTR; the deduced 660-amino acid sequence (R69633) showed an altered C-terminal sequence compared to R69632. Recombinant IL-12 receptor was expressed in COS cells, and can be used for therapeutic or diagnostic purposes.

CC Sequence 660 AA:

Query Match 50.9%; Score 55; DB 1; Length 660;
Best Local Similarity 53.8%; Pred. No. 8.73e+01;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 14 LFLLSRGACACT 26
OY 3 MFLSRNTGEV 15

RESULT 14
ID R69632 standard: Protein: 662 AA.
AC R69632;
DT 05-SEP-1995 (first entry)
DE Human Interleukin-12 receptor.
KW Interleukin-12 receptor; IL-12; immune suppression;
immunosuppressive; graft-versus-host reaction; allograft rejection;
inflammation; autoimmune disease.
OS Homo sapiens.
FH key
FT peptide 1..23 Location/Qualifiers
FT region /label= Sig-peptide
FT /note= "extracellular region"
FT 52..64
FT /note= "sequence motif of cytokine receptor
FT superfamily Cys52..Cys62SW"
FT modified_site 121..123
FT /label= N-linked-glycosylation
FT region 222..226
FT /note= "cytokine receptor superfamily motif
FT (W22SKWS)"
FT modified_site 329..331
FT /label= N-linked-glycosylation
FT modified_site 346..348
FT /label= N-linked-glycosylation
FT modified_site 352..354
FT /label= N-linked-glycosylation
FT modified_site 442..444
FT /label= N-linked-glycosylation
FT modified_site 456..458
FT /label= N-linked-glycosylation
FT region 541..570
FT /label= Transmembrane-region
FT region 571..662
FT /label= Cytoplasmic-tail-region
FT 577..584
FT /label= Cytoplasmic-tail
FT /note= "conserved area of cytoplasmic tail"
FT region 618..629
FT /label= Cytoplasmic-tail
FT /note= "conserved area of cytoplasmic tail"
PN EP-638644-A.
PD 15-FEB-1995.
PF 08-JUL-1994: 110657.
PR 19-JUL-1993: US-094649.
PR 19-JUL-1993: US-094713.
PR 31-MAY-1994: US-248532.
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
PI Chizzonate RA, Chua AO, Gubler UA, Trutt TP;
DR MPI: 95-076349/11.
DR N-PSDB: 083844.
PT DNA encoding a low affinity Interleukin-12 receptor - used to
bind or scavenge IL-12 to cause immune suppression, e.g. to
suppress graft-vs-host reaction, allograft rejection or
PT inflammation, and to treat autoimmune conditions
PS Claim 4: Page 24-27: 61pp: English.
CC A cDNA library of PHA-activated peripheral blood mononuclear cells
CC in vector pEF-BOS was screened for Interleukin-12 (IL-12) receptor
CC cDNAs by panning. An isolated cDNA was sequenced (083844); it
CC encoded a 662-amino acid low affinity IL-12 receptor (R69632).
CC Recombinant IL-12 receptor was expressed in COS cells, and can be
CC used for therapeutic or diagnostic purposes.
SQ Sequence 662 AA;

Query Match 50.9%; Score 55; DB 1; Length 662;
Best Local Similarity 53.8%; Pred. No. 8.73e+01;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 14 LFLSRGACRT 26
OY 3 MFLSRNTGEVRT 15

RESULT 15
ID W12772 standard: Protein: 662 AA.
AC W12772;
DT 12-MAY-1997 (first entry)
DE Human Interleukin-12 beta-1 receptor.
KW Interleukin-12 beta-1 receptor; IL-12; autoimmune disease;
rheumatoid arthritis; inflammatory bowel disease;
multiple sclerosis.
OS Homo sapiens.
FH key
FT peptide 1..20 Location/Qualifiers
FT /label= Sig-peptide
FT /note= "signal peptide cleavage site alternatively
FT follows Ala-23 to Cys-24"
FT domain 24..540
FT /label= Extracellular-domain
FT domain 540..570
FT /label= Transmembrane-domain
FT domain 571..662
FT /label= Cytoplasmic-tail
FT region 52..64
FT /label= Cytokine-receptor-motif
FT /note= "cytokine receptor superfamily motif
FT (Cys52..Cys62SW)"
FT region 222..226
FT /label= Cytokine-receptor-motif
FT /note= "cytokine receptor superfamily motif
FT (W22SKWS)"
FT modified_site 121..123
FT /label= N-linked-glycosylation_site
FT modified_site 329..331
FT /label= N-linked-glycosylation_site
FT modified_site 346..348
FT /label= N-linked-glycosylation_site
FT modified_site 352..354
FT /label= N-linked-glycosylation_site
FT modified_site 442..444
FT /label= N-linked-glycosylation_site
FT modified_site 456..458
FT /label= N-linked-glycosylation_site
PN EP-759466-A2.
PD 26-FEB-1997.
PF 23-JUL-1996: 111807.
PR 01-AUG-1995: US-001701.
PR 30-MAY-1996: US-018674.
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
PI Gubler UA, Presky DH;
DR MPI: 97-147515/14.
DR N-PSDB: T59732.
PT New Interleukin-12 beta-2 receptor and high binding affinity
PT complexes - have a high affinity for Interleukin-12, and are used to
PT treat autoimmune diseases
PS Claim 8: Page 37-41: 53pp: English.
CC Human Interleukin-12 (IL-12) receptor beta-1 receptor (W12772)
CC has a low binding affinity for IL-12, but when complexed with an
CC IL-12 beta-2 receptor (see also W12771), forms a complex with a
CC high binding affinity for IL-12. Its amino acid sequence was
CC deduced from a cDNA clone (T59732) obtd. from human lymphoblasts.
CC IL-12 receptor beta-1 can be expressed on the surface of
CC transformed host cells as a complex with co-expressed IL-12 receptor
CC beta-2, and used in therapeutic compounds. pref. with at least 1
CC cytokine antagonist, to treat autoimmune dysfunctions such as
CC rheumatoid arthritis, inflammatory bowel disease and multiple
CC sclerosis. The receptor protein or complex can also be used to
CC detect (ant)agonists of IL-12 activity.
SQ Sequence 662 AA;

Query Match 50.9%; Score 55; DB 1; Length 662;
Best Local Similarity 53.8%; Pred. No. 8.73e+01;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 14 LFLSRGACRT 26

Mon May 15 08:08:39 2000

US-08-991-628-4a.rag

Page 7

07 :||| :||
3 MFLSRNTGEVRT 15

Search completed: Sat May 13 07:16:47 2000
Job time : 8 secs.

Db 206 TPFLLSRNTEGVRT 220
 Oy 1 TPFLLSRNTEGVRT 15

RESULT 2

ENTRY 1JHUG1 #type complete
 TITLE desmoglein 1 precursor - human
 ALTERNATE_NAMES desmosomal glycoprotein I
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS S16906; A39706; A61254; A61279; S16158

REFERENCE
 #authors Buxton, R.S.
 #submission submitted to the EMBL Data Library, November 1990
 #accession S16906
 #molecule_type mRNA
 #residues 1-1049 ##label BUX

REFERENCE
 #cross-references EMBL:X56554; NID:g30505; PIDN:CAA39976.1; PID:g30506
 #accession A39706
 #authors Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arnemann, J.; Rutman, A.J.; Pidsley, S.C.; Watt, F.M.; Rees, D.A.; Buxton, R.S.; Magee, A.I.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4796-4800
 #title Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.

#cross-references MUID:91271279
 #accession A39706
 #molecule_type mRNA
 #residues 24-1049 ##label WHE

REFERENCE
 #cross-references GB:X5654
 #accession A61254
 #authors Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.
 #journal J. Cell Sci. (1991) 99:809-821
 #title Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.

#cross-references MUID:92121251
 #accession A61254
 #molecule_type mRNA
 #residues 26-1049 ##label NIL

REFERENCE
 #cross-references MUID:92175187
 #accession A61279
 #authors Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee, A.I.
 #journal Biochem. Soc. Trans. (1991) 19:1060-1064
 #title Desmosomal glycoproteins I, II and III: novel members of the cadherin superfamily.

GENETICS
 #status not compared with conceptual translation
 #molecule_type mRNA
 #residues 1-55 ##label WH3

FEATURE
 1-23 #domain signal sequence #status predicted #label SIT
 24-49 #domain propeptide #status predicted #label PRO
 50-1049 #product desmoglein #status predicted #label MAT
 50-548 #domain extracellular #status predicted #label EXT
 52-157 #domain cadherin repeat homology #label CR1
 160-269 #domain cadherin repeat homology #label CR2
 272-385 #domain cadherin repeat homology #label CR3
 392-493 #domain cadherin repeat homology #label CR4
 509-530 #region serine/threonine-rich
 549-569 #domain transmembrane #status predicted #label TMN

572-1049
 840-869 #domain intracellular #status predicted #label INT
 870-899 #domain desmoglein repeat #label DG1
 900-927 #domain desmoglein repeat #label DG2
 928-956 #domain desmoglein repeat #label DG3
 969-1019 #domain desmoglein repeat #label DG4
 110,180 #region glycoprotein repeat #label DG4
 #binding-site carbohydrate (Asn) (covalent) #status predicted

SUMMARY
 #length 1049 #molecular-weight 113715 #checksum 4482

Query Match 81.5%; Score 88; DB 1; Length 1049;
 Best Local Similarity 66.7%; Pred. No. 1,37e-06;
 Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 206 SPMFIRNTEGVRT 220
 Oy 1 TPFLLSRNTEGVRT 15

RESULT 3

ENTRY 1JHUG1 #type complete
 TITLE desmoglein 1 precursor - bovine
 ALTERNATE_NAMES desmoglein BDGM
 ORGANISM #formal_name Bos primigenius taurus #common_name cattle
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS S14603; A38872; A37785; S38721; A48173; S24412

REFERENCE
 #authors Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #submission submitted to the EMBL Data Library, March 1991
 #description Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.

#accession S14603
 #molecule_type mRNA
 #residues 1-1043 ##label KOC

REFERENCE
 #cross-references EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307
 #accession A38872
 #authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1991) 55:200-208
 #title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.

#cross-references MUID:92037656
 #accession A38872
 #molecule_type mRNA
 #residues 1-87,968-1043 ##label KO2

REFERENCE
 #cross-references GB:S64268; GB:S64270
 #accession A37785
 #authors Goodwin, L.; Hill, J.E.; Raynor, K.; Rasz, L.; Menabe, M.; Cowin, P.
 #journal Biochem. Biophys. Res. Commun. (1990) 173:1224-1230
 #title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.

#cross-references MUID:91097553
 #accession A37785
 #molecule_type mRNA
 #residues 44-123, 'V', 125-493 ##label GOO

REFERENCE
 #cross-references GB:M8165; NID:g162966; PIDN:AAA62709.1; PID:g552318
 #accession S38721
 #authors Zimbelmann, R.
 #submission submitted to the EMBL Data Library, February 1991
 #accession S38721
 #molecule_type mRNA
 #residues 44-1043 ##label ZIM

REFERENCE
 #cross-references EMBL:X57784; NID:g436061; PIDN:CAA40930.1; PID:g436062
 #accession A48173
 #authors Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1990) 53:1-12
 #title Identification of desmoglein, a constitutive desmosomal

glycoprotein, as a member of the cadherin family of cell adhesion molecules.
#cross-references MUID:91168965
#accession A48173
#molecule_type mRNA
#residues 44-1001, 'AOPSPAR' ##label KO3
##cross-references GB:X57784
##note this sequence has been revised in references A38872 and S38721

GENETICS
#gene DSG1
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE
1-23
24-49 #domain signal sequence #status predicted #label SIG
50-1043 #domain propeptide #status predicted #label PRO
50-548 #product desmoglein #status predicted #label MAT
52-157 #domain extracellular #status predicted #label EXT
160-269 #domain cadherin repeat homology #label CR1
392-491 #domain cadherin repeat homology #label CR2
372-385 #domain cadherin repeat homology #label CR3
549-574 #domain cadherin repeat homology #label CR4
575-1043 #domain intracellular #status predicted #label TMM
846-875 #domain desmoglein repeat #label DG1
876-905 #domain desmoglein repeat #label DG2
906-933 #domain desmoglein repeat #label DG3
934-962 #domain desmoglein repeat #label DG4
963-1012 #region glycine/serine-rich
110 #binding_site carbohydrate (asn) (covalent) #status experimental
180,496 #binding_site carbohydrate (asn) (covalent) #status predicted
SUMMARY #length 1043 #molecular-weight 112242 #checksum 6897

Query Match 75.0% Score 81; DB 1; Length 1043;
Best Local Similarity 60.0% Pred. No. 5,83e-05;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 206 SPMFLINRYGCIPT 220
1 TPMFLSRNTEGVT 15

RESULT 4
ENTRY S52694 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) ASN1
ALTERNATE_NAMES - yeast (Saccharomyces cerevisiae)
ORGANISM protein P6659.3; Protein YPR145W
#formal_name Saccharomyces cerevisiae
DATE 19-May-1995 #sequence_revision 01-Sep-1995 #text_change 16-Jul-1999
ACCESSIONS S52694; S69033
REFERENCE S52694
#authors Dang, V.D.; Bolotin-Fukuhara, M.; Dalgman-Fornier, B.
#submission submitted to the EMBL Data Library, March 1995
#description Multiple regulatory systems control expression of the Saccharomyces cerevisiae ASN1 and ASN2 genes at the transcription level.
#accession S52694
#molecule_type DNA
#residues 1-572 ##label DAN
##cross-references EMBL:248675; NID:g1163061; PIDN:CAA8594.1; PID:g747902

REFERENCE S69019
#author Fulton, L.
#submission submitted to the EMBL Data Library, November 1995
#description The sequence of S. cerevisiae cosmid 9659.
#accession S69033
#molecule_type DNA
#residues 1-572 ##label FU
##cross-references EMBL:U40829; NID:g1066476; PIDN:AAB68284.1;

PID:g1066479; MIPS:YPR145W

GENETICS
#gene SGD:ASN1
##cross-references SGD:S0006349; MIPS:YPR145W
#map_position 16R
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
FEATURE
2
SUMMARY #active-site Cys (amide transfer) #status predicted
#length 572 #molecular-weight 64470 #checksum 8736

Query Match 60.2% Score 65; DB 2; Length 572;
Best Local Similarity 88.9% Pred. No. 1.74e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 347 TPMFLSRK 355
1 TPMFLSRN 9

RESULT 5
ENTRY S55982 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) ASN2
ALTERNATE_NAMES - yeast (Saccharomyces cerevisiae)
ORGANISM protein G6358; Protein YGR124W
#formal_name Saccharomyces cerevisiae
DATE 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 16-Jul-1999
ACCESSIONS S55982; S64433
REFERENCE S55976
#authors van Dyck, L.; Goffeau, A.
#submission submitted to the EMBL Data Library, December 1994
#description Genes for an asn synthase, a GUG-motif nucleoporin and a putative homeobox-domain protein are identified on a 18.3 kb segment of the yeast chromosome VII also carrying MEPI, PPT1, tree new ORFs, remnants of Ty and three tRNA genes.
#accession S55982
#molecule_type DNA
#residues 1-572 ##label VAN
##cross-references EMBL:X83099; NID:9642340; PIDN:CAA58159.1; PID:g642347

REFERENCE S64428
#authors Van Dyck, L.; Skala, J.; de Mergifosse, P.; Purnelle, B.; Talle, E.; Nawrocki, A.; Del Bino, S.; Goffeau, A.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64433
#molecule_type DNA
#residues 1-572 ##label VAM
##cross-references EMBL:272909; NID:g1323202; PIDN:CAA57135.1; PID:e243463; PID:g1323203; MIPS:YGR124W
##experimental_source strain S288C

GENETICS
#gene SGD:ASN2
##cross-references SGD:S0003356; MIPS:YGR124W
#map_position 7R
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
FEATURE
2
SUMMARY #active-site Cys (amide transfer) #status predicted
#length 572 #molecular-weight 64593 #checksum 9651

Query Match 60.2% Score 65; DB 2; Length 572;
Best Local Similarity 88.9% Pred. No. 1.74e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 346 TPMFLSRK 354
1 TPMFLSRN 9

RESULT 6
ENTRY S33892 #type complete
TITLE hypothetical protein 6 (peti 5' region) - Salmonella

```

ORGANISM      typhimurium
#formal_name Salmonella typhimurium
DATE          19-Mar-1997 #sequence_revision 25-Apr-1997 #text-change
20-Sep-1999

ACCESSIONS   S32892
REFERENCE    S32886
#authors     Friedrich, M.J.; Kinsey, N.E.; Vile, J.; Kadner, R.J.
#journal     Mol. Microbiol. (1993) 8:543-558
#title       Nucleotide sequence of a 13.9kb segment of the 90kb virulence
            Plasmid of Salmonella typhimurium: the presence of fimbrial
            biosynthetic genes.

#cross-references MUID:93316852
#accession     S32892
#status        Preliminary
#molecule_type DNA
#residues      1-295 ##label FRI
#cross-references EMBL:L08613
CLASSIFICATION #superfamily Salmonella typhimurium hypothetical protein 6
                (pepf 5', region)
SUMMARY        #length 295 #molecular-weight 31162 #checksum 8332

Query Match      58.3% Score 63; DB 2; Length 285;
Best Local Similarity 80.0%; Pred. NO. 4,41e-01;
Matches 12; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db      132 TLMFSLSR-TGEVRT 145
        | | | | | | | | |
Oy      1 TFMFLSRNTGEVRT 15

RESULT 7
ENTRY   S56364 #type complete
TITLE   Inner membrane copper tolerance protein cycZ - Escherichia
        coli
ALTERNATE_NAMES thiol:disulfide interchange protein dsbd
ORGANISM #formal_name Escherichia coli
DATE      28-Oct-1995 #sequence_revision 03-Nov-1995 #text-change
21-Nov-1997

ACCESSIONS S56364; I41028; I41037; S42064; F65223; S57220; S47295
REFERENCE   S56314
#authors     Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.;
            Blatter, F.R.
#journal     Nucleic Acids Res. (1995) 23:2105-2119
#title       Analysis of the Escherichia coli genome VI: DNA sequence of
            the region from 92.8 through 100 minutes.
#cross-references MUID:95334362
#accession     S56364
#status        nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues      1-565 ##label BUR
#cross-references EMBL:U14003; NID:g1263172; PID:g536980
#note         The nucleotide sequence was submitted to the EMBL Data
            Library, August 1994

REFERENCE    I41027
#authors     Fong, S.T.; Camakaris, J.; Lee, B.T.
#journal     Mol. Microbiol. (1995) 15:1137-1137
#title       Molecular genetics of a chromosomal locus involved in copper
            tolerance in Escherichia coli K-12.
#cross-references MUID:95349397
#accession     I41028
#status        translated from GB/EMBL/DBJ
#molecule_type DNA
#residues      'M', 78-565 ##label RES
#cross-references EMBL:Z36905; NID:g535290; PID:g581055
#note         In this report, the codon GTG for Val-77 was interpreted
            as a start codon

REFERENCE    I41036
#authors     Crooke, H.; Cole, J.
#journal     Mol. Microbiol. (1995) 15:1139-1150
#title       The biosynthesis of c-type cytochromes in Escherichia coli
            requires a membrane-bound protein, DtpZ, with a protein
            disulphide isomerase-like domain.
#cross-references MUID:95349398

```

```

#accession      I41037
#status         translated from GB/EMBL/DBJ
#molecule_type DNA
#residues       'M', 78-565 ##label RE2
#cross-references EMBL:X77707; NID:9871027; PID:9871029
#note           In this report, the codon GTG for Val-77 was interpreted
                as a start codon

REFERENCE
#authors        S42063
#submission     Crooke, H.R.; Cole, J.A.
#description    Submitted to the EMBL Data Library, February 1994
                The biogenesis of C-type cytochromes in Escherichia coli
                requires an integral membrane protein with a protein
                disulphide isomerase like domain.
#accession      S42064
#molecule_type DNA
#residues       1-328, 'V', 330-565 ##label CRO
#cross-references EMBL:X77707

REFERENCE
#authors        Blatter, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                Kirtpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                Y.
#accession      Science (1997) 277:1453-1462
#title          The complete genome sequence of Escherichia coli K-12.
#cross-references NID:97426617
#accession      F65223
#status         Nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues       1-565 ##label BLAT
#cross-references GB:AE000486; GB:U00096; NID:q1790574; PID:q1790578;
                UMG:pB4136

##experimental-source strain K-12, substrain MG1655

REFERENCE
#authors        Mistakias, D.; Schwager, F.; Raine, S.
#journal        EMBO J. (1995) 14:3435-3424
#title          Identification and characterization of a new disulfide
                isomerase-like protein (DsbD) in Escherichia coli.
#cross-references NID:9534659
#accession      S57220
#status         Preliminary
#molecule_type DNA
#residues       462-540 ##label MIS

GENETICS
#gene           dsbD; cycZ; Cuta2; dlpZ
#keywords       inner membrane; redox-active disulfide; transmembrane protein
#summary        #length 565 #molecular-weight 61795 #checksum 2665

Query Match      58.3% Score 63; DB 2; Length 565;
Best Local Similarity 53.8% Pred. No. 4, 41e-01;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db      373 PVEILLERVIDW 385
Oy      2 PMFLSNTGCV 14

RESULT      8
ENTRY
TITLE       S50482 #type complete
            hypothetical protein YER024w - yeast (Saccharomyces
            cerevisiae)
ORGANISM    #formal_name Saccharomyces cerevisiae
            28-May-1993 #sequence_revision 24-Feb-1995 #text-change
            21-Nov-1997
DATE
ACCESSIONS  S50482
REFERENCE    S50428
            S50428
            Dietrich, F.S.
            submitted to the EMBL Data Library, December 1994
            The sequence of S. cerevisiae cosmids 9537, 9581, 9495, 9867,
            and Lambda clone 5898.
#accession   S50482
#molecule_type DNA
#residues    1-923 ##label DIE

```

```

GENETICS ##cross-references EMBL:U018778; NID:g603592; PID:g603616; MIPS:YER0244
SUMMARY #map_position 5R #length 923 #molecular-weight 103333 #checksum 767

Query Match 58.3%; Score 63; DB 2; Length 923;
Best Local Similarity 53.8%; Pred. No. 4,41e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0

Db 722 MEMLSWVGEIRS 734
Oy 3 MFLLSRNTEVRT 15
|||||:||||:

RESULT 9
ENTRY T08846 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 -
ALTERNATE_NAMES asparagine synthetase 2
ORGANISM #formal_name glycine max #common_name soybean
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
23-Jul-1999
ACCESSIONS T08846
REFERENCE 216490
#authors Hughes, C.A.; Beard, H.; Matthews, B.F.
#journal Plant Mol. Biol. (1997) 33:301-311
#title Molecular cloning and expression of two, cDNAs encoding
asparagine synthetase in soybean.
#accession T08846
#molecule_type mRNA
#residues 1-581 #label HUG
#cross-references EMBL:U77678; NID:g17783370
#experimental_source strain Century; leaf
GENETICS AS2
#gene
#description catalyzes ATP-dependent transfer of the amide group of
glutamine to aspartate generating asparagine and glutamate
#pathway asparagine biosynthesis
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
SUMMARY #length 581 #molecular_weight 65609 #checksum 7265

Query Match 56.5%; Score 61; DB 2; Length 581;
Best Local Similarity 77.8%; Pred. No. 1.10e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPFMLSRK 330
Oy 1 TPFMLSRN 9
|||||:||||:

RESULT 10
ENTRY AJPMN2 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 -
ALTERNATE_NAMES asparagine synthetase (glutamine-hydrolyzing)
ORGANISM #formal_name Pisum sativum #common_name garden pea
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
18-Jun-1999
ACCESSIONS S11443
REFERENCE S11443
#authors Tsai, F.Y.; Coruzzi, G.M.
#journal EMBO J. (1990) 9:323-332
#title Dark-induced and organ-specific expression of two asparagine
synthetase genes in Pisum sativum.
#cross-references MUID:90151604
#accession S11443
#molecule_type DNA
#residues 1-583 #label TSA
#cross-references EMBL:X52180; NID:g20651; PIDN:CA56430.1; PID:g20652
GENETICS

```

```

#gene AS2
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
FEATURE
2
SUMMARY #active-site Cys (amide transfer) #status predicted
#length 583 #molecular-weight 65649 #checksum 2836

Query Match 56.5%: Score 61: DB 1: Length 583:
Best Local Similarity 77.8%: Pred. No. 1.10e+00:
Matches 7: Conservative 2: Mismatches 0: Indels 0: Caps 0:

DB 322 TPMFLMSRK 330
QY 1 TPMFLSRN 9
|||||:
|
RESULT 11
ENTRY T12989 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) -
Arabidopsis thaliana
ORGANISM asparagine synthetase, glutamine-dependent; protein T21L8_90
ORGANISM $format_name Arabidopsis thaliana $common_name mouse-ear
cress
DATE 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change
20-Sep-1999
ACCESSIONS T12989
REFERENCE T12986
#authors Choisme, N.; Robert, C.; Brotlier, P.; Wincker, P.;
Cattolico, L.; Artiguenave, F.; Saulin, W.; Weissenbach,
J.; Meyers, H.W.; Mayer, K.F.X.; Lemcke, K.; Schuebler, C.;
Queller, F.; Selaouat, M.
Submitted to the Protein Sequence Database, July 1999
T12989
#submission submitted to the Protein Sequence Database, July 1999
#accession T12989
#status Preliminary
#molecule_type DNA
##residues 1-584 ##label CHO
##cross-references EMBL:AL096860; GSPDB:GN00061; ATSP:T21L8_90
##experimental_source cultivar Columbia; BAC clone T21L8
GENETICS
#gene ATSP:T21L8_90
#map_position 3
#introns 27/2: 73/3: 105/3: 216/3: 270/3: 371/3: 416/3: 508/3
KEYWORDS ligase
SUMMARY #length 584 #molecular-weight 65620 #checksum 4444

Query Match 56.5%: Score 61: DB 2: Length 584:
Best Local Similarity 77.8%: Pred. No. 1.10e+00:
Matches 7: Conservative 2: Mismatches 0: Indels 0: Caps 0:

DB 322 TPMFLMSRK 330
QY 1 TPMFLSRN 9
|||||:
|
RESULT 12
ENTRY S69182 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 1 -
Lotus japonicus
ORGANISM #formal_name Lotus japonicus
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
16-Jul-1999
ACCESSIONS S69182; S57931
REFERENCE S69182
#authors Waterhouse, R.N.; Smyth, A.J.; Massonneau, A.; Prosser, I.M.;
Clarkson, D.T.
Plant Mol. Biol. (1996) 30:883-897
Molecular cloning and characterisation of asparagine
synthetase from Lotus japonicus: dynamics of asparagine
synthesis in N-sufficient conditions.
#cross-references MORD:96270368
#accession S69182
#status nucleic acid sequence not shown
#molecule_type mRNA

```

```

##residues      1-586 ##label WAT
##cross-references EMBL:X89409; NID:g897770; PIDN:CMA61589.1;
                      PID:g897771
##experimental_source strain B-129

GENETICS
#gene          AS1
#CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
#KEYWORDS       asparagine biosynthesis; ligase
#SUMMARY        length 586 #molecular-weight 66461 #checksum 3372

Query Match           56.5%; Score 61; DB 2; Length 586;
Best Local Similarity 77.8%; Pred. NO. 1.10e+00;
Matches              7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db             322 TPMFLMSRK 330
Oy             1 TPMFLMSRN 9

RESULT         13
ENTRY          S69183          #type complete
TITLE          asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 -
ORGANISM       Lotus japonicus
DATE           19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
               16-Jul-1999
ACCESSIONS     S69183; S57932
REFERENCE      S69182
#authors       Waterhouse, R.N.; Smyth, A.J.; Massonieu, A.; Prosser, I.M.
#journal       Plant Mol. Biol. (1996) 30:883-897
#title         Molecular cloning and characterisation of asparagine
               synthetase from Lotus japonicus: dynamics of asparagine
               synthesis in N-sufficient conditions.
#cross-references MUID:96270368
#accession     S69183
#status        nucleic acid sequence not shown
#molecule_type mRNA
#residues      1-586 ##label WAT
#cross-references EMBL:X89410; NID:g897772; PIDN:CMA61590.1;
               PID:g897773
##experimental_source strain B-129

GENETICS
#gene          AS2
#CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
#KEYWORDS       asparagine biosynthesis; ligase
#SUMMARY        length 586 #molecular-weight 65969 #checksum 518

Query Match           56.5%; Score 61; DB 2; Length 586;
Best Local Similarity 77.8%; Pred. NO. 1.10e+00;
Matches              7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db             322 TPMFLMSRK 330
Oy             1 TPMFLMSRN 9

RESULT         14
ENTRY          AJPNM1          #type complete
TITLE          asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 1 -
               garden pea
ALTERNATE_NAMES asparagine synthetase (glutamine-hydrolyzing)
ORGANISM         #formal name Pisum sativum #common_name garden pea
DATE            31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
               18-Jun-1999
ACCESSIONS     S11444
SOURCE         Tsai, F.Y.; Cornuzzi, G.M.
AUTHORS        EMBO J. (1990) 9:323-332
JOURNAL        Dark-induced and organ-specific expression of two asparagine
               synthetase genes in Pisum sativum.
CROSS-REFERENCES MUID:90151604
ACCESSION     S11444

```

```

##molecule_type DNA
##residues 1-586 ##label TSA
##cross-references EMBL:X52179; NID:g20649; PIDN:CAA36429.1; PID:g20650
COMMENT This protein is one of a family of glutamine amidotransferases that
have dual specificity to utilize either glutamine or ammonia as a
substrate. Two functional domains have been identified: an
aminator domain, which catalyzes the ammonia-dependent reaction,
and a glutamine amide transfer domain, which binds glutamine and
transfers the amide to the aminator domain.
GENETICS
#gene AS1
#classification #superfamily asparagine synthase (glutamine-hydrolyzing)
#keywords asparagine biosynthesis; ligase
FEATURE
175-586
2 #domain aminator #label AMN
#active-site Cys (amide transfer) #status predicted
#length 586 #molecular-weight 66353 #checksum 3350
SUMMARY
Query Match 56.5%; Score 61; DB 1; Length 586;
Best Local Similarity 77.8%; Pred. No. 1,10e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 323 TPMFLMSRK 331
|||||:
1 TPMFLSRN 9
RESULT 15
ENTRY T02978 #type complete
TITLE asparagine synthase (EC 6.3.-.-) - maize
ALTERNATE_NAMES asparagine synthetase
ORGANISM #formal_name zea mays #common_name maize
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text-change
16-Jul-1999
ACCESSIONS T02978
REFERENCE 214805
#authors Chevalier, C.; Bourgeois, E.; Just, D.; Raymond, P.
#journal Plant J. (1996) 9:1-11
#title Metabolic regulation of asparagine synthetase gene expression
in maize (Zea mays L.) root tips.
#cross-references MUID:96158342
#accession T02978
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
##residues 1-586 ##label CHE
##cross-references EMBL:X82849; NID:g984261; PIDN:CAA58052.1;
PID:g984262
##experimental_source cultivar var.DEA; root meristem
GENETICS
#note AS
#classification #superfamily asparagine synthase (glutamine-hydrolyzing)
#keywords ligase
SUMMARY #length 586 #molecular-weight 66577 #checksum 5815
Query Match 56.5%; Score 61; DB 2; Length 586;
Best Local Similarity 77.8%; Pred. No. 1,10e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 322 TPMFLMSRK 330
|||||:
1 TPMFLSRN 9
Search completed: Sat May 13 07:16:22 2000
Job time : 8 secs.

```

M P S R E I H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:11:13 2000; Maspar time 80.16 Seconds
Tabular output not generated. 5.699 Million cell updates/sec

Title: >US-08-991-628-4
Description: (1-15) from US08991628.pep
Perfect Score: 108
Sequence: 1 TPMFLSNTGTVRT 15

Scoring table: PAM 150
Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 27.727; Variance 32.559; scale 0.852

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|-------------|-----------|
| 1 | 108 | 100.0 | 999 | 1 | DSG3_HUMAN | 4.18e-13 |
| 2 | 88 | 81.5 | 1049 | 1 | DSG1_HUMAN | 1.09e-07 |
| 3 | 81 | 75.0 | 1043 | 1 | DSG1_BOVIN | 6.76e-06 |
| 4 | 65 | 60.2 | 571 | 1 | ASN2_YEAST | 4.34e-02 |
| 5 | 65 | 60.2 | 571 | 1 | ASN1_YEAST | 4.34e-02 |
| 6 | 63 | 58.3 | 488 | 1 | DSBD_ECOLI | 1.20e-01 |
| 7 | 63 | 58.3 | 923 | 1 | YEL4_YEAST | 1.20e-01 |
| 8 | 61 | 56.5 | 524 | 1 | ASN5_SANAU | 3.25e-01 |
| 9 | 61 | 56.5 | 582 | 1 | ASN2_PEA | 3.25e-01 |
| 10 | 61 | 56.5 | 583 | 1 | ASN2_ARATH | 3.25e-01 |
| 11 | 61 | 56.5 | 585 | 1 | ASN2_TRIVS | 3.25e-01 |
| 12 | 61 | 56.5 | 585 | 1 | ASN2_LOTJA | 3.25e-01 |
| 13 | 61 | 56.5 | 585 | 1 | ASN1_PEA | 3.25e-01 |
| 14 | 61 | 56.5 | 585 | 1 | ASN5_MAIZE | 3.25e-01 |
| 15 | 61 | 56.5 | 585 | 1 | ASN5_BRAOL | 3.25e-01 |
| 16 | 61 | 56.5 | 585 | 1 | ASN1_LOTJA | 3.25e-01 |
| 17 | 61 | 56.5 | 590 | 1 | ASN5_ORYSA | 3.25e-01 |
| 18 | 59 | 54.6 | 589 | 1 | ASN5_ASPOF | 8.60e-01 |
| 19 | 59 | 54.6 | 1103 | 1 | CYGD_HUMAN | 8.60e-01 |
| 20 | 57 | 52.8 | 554 | 1 | WH01_VARY | 2.22e+00 |
| 21 | 57 | 52.8 | 579 | 1 | DSBD_ECOLI | 2.22e+00 |
| 22 | 57 | 52.8 | 1117 | 1 | DSG2_HUMAN | 2.22e+00 |
| 23 | 57 | 52.8 | 1117 | 1 | DSG2_HUMAN | 2.22e+00 |

| | | | | | | | |
|----|----|------|------|---|-------------|------------------------|----------|
| 24 | 55 | 50.9 | 124 | 1 | YM06_MYCTU | HYPOTHETICAL 14.0 KD P | 5.59e+00 |
| 25 | 55 | 50.9 | 186 | 1 | ATPD_RHOBL | ATP SYNTHASE DELTA CHA | 5.59e+00 |
| 26 | 55 | 50.9 | 375 | 1 | CISY_MYCSM | CITRATE SYNTHASE (EC 4 | 5.59e+00 |
| 27 | 55 | 50.9 | 662 | 1 | I12R_HUMAN | INTERLEUKIN-12 RECEPT | 5.59e+00 |
| 28 | 54 | 50.0 | 152 | 1 | VE6_HPVO3 | E6 PROTEIN. | 8.79e+00 |
| 29 | 54 | 50.0 | 244 | 1 | ANR_PSEAE | TRANSCRIPTIONAL ACTIVA | 8.79e+00 |
| 30 | 54 | 50.0 | 244 | 1 | FNRA_PSEST | TRANSCRIPTIONAL ACTIVA | 8.79e+00 |
| 31 | 54 | 50.0 | 325 | 1 | MC5R_MOUSE | MELANOCORTIN-5 RECEPT | 8.79e+00 |
| 32 | 54 | 50.0 | 325 | 1 | MC5R_RAT | MELANOCORTIN-5 RECEPT | 8.79e+00 |
| 33 | 54 | 50.0 | 379 | 1 | FD3E_TOBAC | OMEGA-3 FATTY ACID DES | 8.79e+00 |
| 34 | 54 | 50.0 | 423 | 1 | CGB1_RAT | G2/MITOTIC-SPECIFIC CY | 8.79e+00 |
| 35 | 54 | 50.0 | 429 | 1 | CGB1_MESAU | G2/MITOTIC-SPECIFIC CY | 8.79e+00 |
| 36 | 54 | 50.0 | 433 | 1 | CGB1_HUMAN | G2/MITOTIC-SPECIFIC CY | 8.79e+00 |
| 37 | 54 | 50.0 | 582 | 1 | YC73_MYCTU | HYPOTHETICAL ABC TRANS | 8.79e+00 |
| 38 | 54 | 50.0 | 1538 | 1 | LHR_ECOLI | PROBABLE ATP-DEPENDENT | 8.79e+00 |
| 39 | 53 | 49.1 | 171 | 1 | WH01_RACVI | PROTEIN-TYROSINE PHOSP | 1.37e-01 |
| 40 | 53 | 49.1 | 325 | 1 | MC5R_BOVIN | MELANOCORTIN-5 RECEPT | 1.37e-01 |
| 41 | 53 | 49.1 | 713 | 1 | KP78_HUMAN | PURATIVE SERINE/THREON | 1.37e-01 |
| 42 | 53 | 49.1 | 789 | 1 | VIRB4_AGR75 | VIRB4 PROTEIN PRECURSO | 1.37e-01 |
| 43 | 53 | 49.1 | 789 | 1 | VIRB4_AGR75 | VIRB4 PROTEIN PRECURSO | 1.37e-01 |
| 44 | 53 | 49.1 | 943 | 1 | YIM5_CABEL | HYPOTHETICAL 105.9 KD | 1.37e-01 |
| 45 | 53 | 49.1 | 1220 | 1 | PTC1_BRAPE | PATCHED PROTEIN HOMOLO | 1.37e-01 |

ALIGNMENTS

RESULT 1
ID DSG3_HUMAN STANDARD: PRT: 999 AA.

AC P32926:

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).

GN DSG3.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Homnidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 92069753.

RA Anagel M., Klaus-Kovtun V., Stanley J.R.;

RT "Autoantibodies against a novel epithelial cadherin in pemphigus

RT vulgaris, a disease of cell adhesion.";

RL Cell 67:869-877(1991)

CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSE JUNCTIONS.

CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE

CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND

CC CARCINOMAS.

CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS

CC (POTENTIAL).

CC -1- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN

CC DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE

CC LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES

CC AGAINST DSG3.

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSE SUBFAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M76482; AAA60230.1; -

DR PIR: A41088; IYHUG3.

DR HSSP: P09803; 1EDH.

DR MIM: 169615; -

DR PROSITE: PS00232; CADHERIN; 3.

DR PFM: PF00028; cadherin; 4.

DR Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;

KW

```
KW Calcium-binding; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 49 POTENTIAL.
FT CHAIN 50 999 DESMOGLEIN 3.
FT DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 616 640 POTENTIAL.
FT DOMAIN 641 999 CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 158 CADHERIN 1.
FT REPEAT 159 268 CADHERIN 2.
FT REPEAT 269 383 CADHERIN 3.
FT REPEAT 386 499 CADHERIN 4.
FT REPEAT 910 935 DESMOGLEIN REPEAT 1.
FT REPEAT 936 966 DESMOGLEIN REPEAT 2.
FT CARBOHYD 110 110 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 459 459 POTENTIAL.
FT CARBOHYD 545 545 POTENTIAL.
SQ SEQUENCE 999 AA; 107503 MM; 60479DD46AC219A1 CRC64;

Query Match 100.0%; Score 108; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 4.18e-13;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 206 TPFLLSRNTGEVRT 220
OY 1 TPFLLSRNTGEVRT 15

RESULT 2
ID DSG1-HUMAN STANDARD: PRT; 1049 AA.
AC 002413;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGL).
GN DSG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KERATINOCYTES;
RX Wheeler G.N., Parker A.E., Thomas C.L., Ataliotis P., Poynter D.,
  Arreman J., Rutman A.J., Pidsley S.C., Watt F.M., Rees D.A.,
  Buxton R.S., Magee A.I.;
  "Desmosomal glycoprotein DGL, a component of intercellular desmosome
  junctions, is related to the cadherin family of cell adhesion
  molecules.";
  Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).
RT FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME
  INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
  FILAMENTS MEDIATING CELL-CELL ADHESION
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL AND ESOPHAGUS.
CC -1- DOMAIN: CALCULIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
  (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
DR EMBL; X56654; CA39976.1; -
DR PIR; S16906; IJHUG1.
DR HSSP; P09803; IEDH.
DR MIM; 125670; -
DR PROSITE; PS00232; CADHERIN. 2.
DR PFM; PF00028; cadherin. 4.
```

```
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 49 POTENTIAL.
FT CHAIN 50 1049 DESMOGLEIN 1.
FT DOMAIN 50 545 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 546 570 POTENTIAL.
FT DOMAIN 571 1049 CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 158 CADHERIN 1.
FT REPEAT 159 270 CADHERIN 2.
FT REPEAT 271 385 CADHERIN 3.
FT REPEAT 386 497 CADHERIN 4.
FT REPEAT 813 839 DESMOGLEIN REPEAT 1.
FT REPEAT 840 869 DESMOGLEIN REPEAT 2.
FT REPEAT 870 899 DESMOGLEIN REPEAT 3.
FT REPEAT 900 927 DESMOGLEIN REPEAT 4.
FT REPEAT 928 956 DESMOGLEIN REPEAT 5.
FT DOMAIN 969 1019 GLY/SER-RICH.
FT CARBOHYD 36 36 POTENTIAL.
FT CARBOHYD 110 110 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
SQ SEQUENCE 1049 AA; 113715 MM; EEEL25655B9D6619 CRC64;

Query Match 81.5%; Score 88; DB 1; Length 1049;
Best Local Similarity 66.7%; Pred. No. 1.09e-07;
Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 206 SPMFLSRNTGEVRT 220
OY 1 TPFLLSRNTGEVRT 15
```

```
RESULT 3
ID DSG1-BOVIN STANDARD: PRT; 1043 AA.
AC 003763;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGL).
GN DSG1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-MUZZLE EPITHELIUM;
RX Koch P.J., Goldschmidt M.D., Zimbelmann R., Franke W.W.;
  Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
RT SEQUENCE OF 44-1043 FROM N.A.
CC TISSUE-MUZZLE EPITHELIUM;
CC MEDLINE; 9168965.
RA Koch P.J., Walsh M.J., Schmelz M., Goldschmidt M.D.,
  Zimbelmann R., Franke W.W.;
  "Identification of desmoglein, a constitutive desmosomal
  glycoprotein, as a member of the cadherin family of cell adhesion
  molecules.";
  Eur. J. Cell Biol. 53:1-12(1990).
RN [3]
RP REVISIONS, AND SEQUENCE OF 101-123.
RX MEDLINE; 92037656.
RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R.,
  Franke W.W.;
  "Complete amino acid sequence of the epidermal desmoglein precursor
  polypeptide and identification of a second type of desmoglein gene.";
  Eur. J. Cell Biol. 55:200-208(1991).
RN [4]
RP SEQUENCE OF 44-493 FROM N.A.
RX MEDLINE; 91097553.
RA Goodwin L., Hill J.E., Raynor K., Raszi L., Manabe M., Cowin P.;
  "Desmoglein shows extensive homology to the cadherin family of cell
  adhesion molecules.";
```

```

RL Biochem. Biophys. Res. Commun. 173:1224-1230(1990).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58466; CAA41380.1; -
DR EMBL: X57784; CAA40930.1; -
DR EMBL: M58165; AAA62709.1; -
DR PIR: S14603; IJBOG1.
DR HSSP: P09803; IEDH.
DR PROSITE: P500232; CADHERIN; 2.
DR PFAM: PF00028; cadherin; 3.
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
FT SIGNAL 1 23
FT PROPEP 24 49
FT CHAIN 50 1043
FT DOMAIN 50 548
FT TRANSMEM 549 573
FT DOMAIN 574 1043
FT REPEAT 50 158
FT REPEAT 159 270
FT REPEAT 271 385
FT REPEAT 386 498
FT REPEAT 499 845
FT REPEAT 846 875
FT REPEAT 876 905
FT REPEAT 906 933
FT REPEAT 934 962
FT DOMAIN 963 1012
FT CARBOHYD 110 110
FT CARBOHYD 180 180
FT CARBOHYD 486 486
FT CONFLICT 124 124
SQ SEQUENCE 1043 AA; 112243 MW; ADE46133F8B7C11 CRC64;

Query Match
Best Local Similarity 75.0%; Score 81; DB 1; Length 1043;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 206 SPMEIINRTGEIRT 220
OY 1 TPMEFLSRNTGEVRT 15

RESULT 4
ID ASN2_YEAST STANDARD: PRT: 571 AA.
AC P49090.
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)
DT 01-NOV-1997 (Rel. 35; Last annotation update)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 2 (EC 6.3.5.4)
DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 2).
DE ASN2 OR YGR124W OR G6356.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / F11679;

```

```

RX MEDLINE: 97197982.
RA van Dyck L., Tettelin H., Purnelle B., Goffeau A.;
RT An 18.3 kb DNA fragment from yeast chromosome VII carries four
RT unknown open reading frames, the gene for an Asn synthase, remnants
RT of Ty and three tRNA genes.
RL yeast 13:171-176(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA van Dyck L., Skala J., de Wergifosse P., Purnelle B., Talla E.,
RA Nawrocki A., del Bino S., Goffeau A.;
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARATE + L-GLUTAMINE -> AMP +
CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X83099; CAA58159.1; -
DR EMBL: 272909; CAA97135.1; -
DR SGD: L0003156; ASN2.
DR PROSITE: P500443; GATASE_TYPE_II; 1.
DR PFAM: PF00310; GATase_2; 1.
DR PFAM: PF00733; Asn_synthase; 1.
KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
KW Multigene family.
FT INIT_MER 0 0
FT ACT_SITE 1 1
FT ACT_SITE 1 1
SQ SEQUENCE 571 AA; 64461 MW; 086D03EC77E6F00A CRC64;

Query Match
Best Local Similarity 88.9%; Score 65; DB 1; Length 571;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 345 TPMEFLSRK 353
OY 1 TPMEFLSRN 9

RESULT 5
ID ASN1_YEAST STANDARD: PRT: 571 AA.
AC P49089.
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)
DT 15-FEB-2000 (Rel. 39; Last annotation update)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 1 (EC 6.3.5.4)
DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 1).
DE ASN1 OR YPR145W OR P9659.3
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA Dang V.D., Valens M., Bolotin-Fukuhara M., Daisman-Fornier B.;
RA MEDLINE: 97109535.
RT Cloning of the ASN1 and ASN2 genes encoding asparagine synthetases
RT in Saccharomyces cerevisiae: differential regulation by the
RT CCAAT-box-binding factor.
RL Mol. Microbiol. 22:681-692(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,

```



```

Matches      7; Conservative      3; Mismatches      3; Indels      0; Gaps      0;
Db          296 PFYLLERVIQDWM 308
              ||||| | ||| :
Oy           2 PFYLLSRNTGEVTR 14

RESULT       7
ID            YELLA_YEAST        STANDARD:      PRT:      923 AA.
AC            P40017             STANDARD:      PRT:
DT            01-FEB-1995 (Rel. 31, Created)
DT            01-FEB-1995 (Rel. 31, Last sequence update)
DT            01-FEB-1995 (Rel. 31, Last annotation update)
DE            HYPOTHEICAL 103.3 KD PROTEIN IN PRO3-GCD1 INTERGENIC REGION.
GN            YER024W.
OS            Saccharomyces cerevisiae (Baker's yeast).
OC            Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
NC            Saccharomycetaceae; Saccharomyces.
RN            (1)
RP            SEQUENCE FROM N.A.
RC            STRAIN-S288C / AB972;
RA            Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA            Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA            Chung E., Duncan M., Guzman E., Hartwell G., Hunicker-Smith S.,
RA            Hyman R., Kayser A., Komp C., Laskerl D., Lew H., Lin D.,
RA            Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA            Petzel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA            Taylor P., Wei Y., Yelton M., Botstein D., Davis R.M.;
RL            Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC            -I- SIMILARITY: BELONGS TO THE CARNITINE/CHOLINE ACETYLTRANSFERASE
CC            FAMILY.
CC            -----
CC            This SWISS-PROT entry is copyright. It is produced through a collaboration
CC            between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC            the European Bioinformatics Institute. There are no restrictions on its
CC            use by non-profit institutions as long as its content is in no way
CC            modified and this statement is not removed. Usage by and for commercial
CC            entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC            or send an email to license@isb-sib.ch).
CC            -----
DR            DR EMBL; U180778; AAB64557.1; '-'
DR            PROSITE; PS00439; ACYLTRANSF_C_1; FALSE_NEG.
DR            PROSITE; PS00440; ACYLTRANSF_C_2; 1.
KW            Pfam; PF00755; Carn_acylttransf.1.
KW            Hypothetical protein; transferase; Acyltransferase.
SQ            SEQUENCE 923 AA; 103333 MW; B59A881491D68A7 CRC64;

Query March      58.3%; Score 63; DB 1; Length 923;
Best Local Similarity 53.8%; Pred. No. 1,20e+01;
Matches      7; Conservative      4; Mismatches      2; Indels      0; Gaps      0;
Db          722 MFMSUWVGIEIRS 734
              ||||| | ||| :
Oy           3 MFLLSRNTGEVRT 15

RESULT       8
ID            ASNS_SAMAU         STANDARD:      PRT:      524 AA.
AC            Q24338;
DT            15-DEC-1998 (Rel. 37, Created)
DT            15-DEC-1998 (Rel. 37, Last sequence update)
DT            15-DEC-1998 (Rel. 37, Last annotation update)
DE            ASPARGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
DE            DEPENDENT ASPARGINE SYNTHETASE).
GN            AND1.
OS            Sandersonia aurantiaca (Christmas-bells) (Chinese-lantern lily).
OC            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC            eumhyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Liliales;
OC            Liliaceae; Sandersonia.
RN            (1)
RP            SEQUENCE FROM N.A.
RC            TISSUE-TEMPAL;
RA            Eason J.R., King G.A.;
```

RT "Nucleoside sequence of cDNA encoding asparagine synthetase from *Sandersonia aurantiaca*."

RL (11) Plant Gene Register PGR97-112.

CC -1- FUNCTION: COULD PLAY A ROLE IN REMOBILIZATION OF NITROGEN IN FLOWERS DURING SENESCENCE.

CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP + PHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.

CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.

CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE AMIDOTRANSFERASES.

CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.sib.ch/announce/or_send_an_email_to_license@isb.sib.ch).

CC -----

DR EMBL: AF005724; AAB71532.1; .

DR PROSITE: PS00443; GATASE_TYPE_II. 1.

DR Pfam: PF00310; GATase_2; 1.

DR Pfam: PF00733; Asn-synthase; 1.

DR Lgase: Asparagine biosynthesis; Glutamine amidotransferase.

FT INT_MET 0 BY SIMILARITY.

FT ACT_SITE 1 1 GATASE (BY SIMILARITY).

SO SEQUENCE 524 AA; 59477 MW; 5C74710231CFF27B CRC64;

Query Match 56.5%; Score 61; DB: 1; Length 524;
Best Local Similarity 77.8%; Pred. No. 3,25e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 321 TPMSLNRK 329
QY 1 TPMSLNRN 9
|||||:

RESULT 9
ID ASN2_PEA STANDARD; PRT; 582 AA.
AC P19252; O49926;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ASPARAGINE SYNTHETASE, ROOT [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4)
DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).
GN AS2.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae;
OC Pisum.
OC [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-CV, SPARKLE; TISSUE=ROOT;
RX MEDLINE: 90151604.
RA Tsai F.Y., Coruzzi G.M.;
RT "Dark-induced and organ-specific expression of two asparagine synthetase genes in *Pisum sativum*.";
RL EMO J. 9:323-332(1990).
RN [2]
RP SEQUENCE OF 1-83 FROM N.A.
RC STRAIN-CV, FELTHAM FIRST;
RA Ngai N., Tsai F.Y., Coruzzi G.M.;
RT "Light-induced transcriptional repression of the pea *Asi* gene: identification of cis-elements and trans factors.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARATE + L-GLUTAMINE -> AMP + PHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- TISSUE SPECIFICITY: ROOTS.
CC -1- INDUCTION: DARK-INDUCED.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE

CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52180; CAA36430.1; -
CC EMBL: Y13322; CAA73763.1; -
CC PIR: S11443; ASPMN2.
CC DR PROSITE: PS00443; GATASE_TYPE_II; 1.
CC DR PFAM: PF00310; GATase_2; 1.
CC DR PFAM: PF00733; Asn_synthase; 1.
CC LIGase: Asparagine biosynthesis; glutamine amidotransferase;
CC Multigene family.
CC KW INIT_MER 0 BY SIMILARITY.
CC FT ACT_SITE 1 GATASE (BY SIMILARITY).
CC FT CONFLICT 33 E -> D (IN REF. 2).
CC FT CONFLICT 44 Y -> F (IN REF. 2).
CC FT CONFLICT 47 Q -> H (IN REF. 2).
CC FT CONFLICT 77 I -> L (IN REF. 2).
CC FT CONFLICT 77 L (IN REF. 2).
CC SQ SEQUENCE 582 AA; 65518 MW; F98546B4B3456265 CRC64;
DB 321 TPWFLLSRK 329
OY 1 TPWFLLSRN 9
Query Match 56.5%; Score 61; DB 1; Length 582;
Best Local Similarity 77.8%; Pred. No. 3,25e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
RESULT 10
ID ASN_ARATH STANDARD; PRT; 583 AA.
AC P49078;
DT 01-FEB-1996 (rel. 33, Created)
DT 01-FEB-1996 (rel. 33, Last sequence update)
DT 01-FEB-1996 (rel. 33, Last annotation update)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
DE DEPENDENT ASPARAGINE SYNTHETASE).
GN ASN1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
OC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-CV. COLUMBIA;
RX MEDLINE: 95148732.
RA Lam H.M., Peng S.S., Cortuzzi G.M.;
RT "Metabolic regulation of the gene encoding glutamine-dependent
RT asparagine synthetase in Arabidopsis thaliana."
RT Plant Physiol. 106:1347-1357(1994).
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE - AMP +
CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: L29083; AAA74359.1; -
DR HSSP: P17169; IGMS.
DR PROSITE: PS00443; GATASE_TYPE_II; 1.
DR PFAM: PF00310; GATase_2; 1.
DR PFAM: PF00733; Asn_synthase; 1.
CC LIGase: Asparagine biosynthesis; glutamine amidotransferase;
CC Multigene family.
CC KW INIT_MER 0 BY SIMILARITY.
CC FT ACT_SITE 1 GATASE (BY SIMILARITY).
CC FT ACT_SITE 1 GATASE (BY SIMILARITY).
CC SQ SEQUENCE 583 AA; 65489 MW; C2C257781D8C7D23 CRC64;
DB 321 TPWFLLSRK 329
OY 1 TPWFLLSRN 9
Query Match 56.5%; Score 61; DB 1; Length 583;
Best Local Similarity 77.8%; Pred. No. 3,25e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
RESULT 11
ID ASN_TRIVS STANDARD; PRT; 585 AA.
AC O24661;
DT 15-DEC-1998 (rel. 37, Created)
DT 15-DEC-1998 (rel. 37, Last sequence update)
DT 15-FEB-2000 (rel. 39, Last annotation update)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
DE DEPENDENT ASPARAGINE SYNTHETASE).
GN AS.
OS Trichysaria versicolor.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Scrophulariaceae;
OC Tripsacra.
OC [1]
RN SEQUENCE FROM N.A.
RP TISSUE-ROOT;
RX MEDLINE: 99051324.
RA Delavault P., Estabrook E., Albrecht H., Wrobel R., Yoder J.I.;
RT "Host-root exudates increase gene expression of asparagine synthetase
RT in the roots of a hemiparasitic plant triphysaria versicolor
RT (Scrophulariaceae)."
RT Gene 222:155-162(1998).
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE - AMP +
CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF014055; AAD5033.1; -
DR EMBL: AF014056; AAD5034.1; -
DR EMBL: AF014057; AAD5035.1; -
DR PROSITE: PS00443; GATASE_TYPE_II; 1.
DR PFAM: PF00310; GATase_2; 1.
DR PFAM: PF00733; Asn_synthase; 1.
CC LIGase: Asparagine biosynthesis; glutamine amidotransferase;
CC KW INIT_MER 0 BY SIMILARITY.
CC FT ACT_SITE 1 GATASE (BY SIMILARITY).
CC FT ACT_SITE 1 GATASE (BY SIMILARITY).
CC SQ SEQUENCE 585 AA; 65560 MW; FAF2E622912117B5 CRC64;
DB 321 TPWFLLSRK 329
OY 1 TPWFLLSRN 9
Query Match 56.5%; Score 61; DB 1; Length 585;
Best Local Similarity 77.8%; Pred. No. 3,25e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 321 TPEFLMSRK 329
 QY 1 TPEFLMSRN 9

RESULT 12
 ID ASN2 LOTJA STANDARD: PRT: 585 AA.
 AC P49093:
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 2 (EC 6.3.5.4)
 DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 2).
 GN AS2.
 OS Lotus Japonicus.
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
 OC core eudicots: Rosidae: eurosids I: Fabales; Fabaceae; Papilionoideae;
 OC Lotus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. GIFU / B-129;
 RX MEDLINE: 96370368.
 RA Waterhouse R.N., Smyth A.J., Massoneau A., Prosser I.M.,
 RA Clarkson D.T.;
 RT "Molecular cloning and characterisation of asparagine synthetase from
 RT Lotus Japonicus: dynamics of asparagine synthesis in N-sufficient
 RT conditions.";
 RT Plant Mol. Biol. 30:883-897(1996).
 RL -1- CATALYTIC ACTIVITY: ATP + L-ASPARATE + L-GLUTAMINE -> AMP +
 CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
 CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
 CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X89410; CAA61590.1; -
 DR HSSP: P17169; IGMS.
 DR PROSITE: PS00443; GATASE_TYPE_II; 1.
 DR PFAM: PF00310; GATase_2; 1.
 DR PFAM: PF00733; Asn_synthase; 1.
 KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
 KM Multigene family.
 FT INIT_MET 0 BY SIMILARITY.
 FT ACT_SITE 1 GATASE (BY SIMILARITY).
 FT SITE 1 GATASE (BY SIMILARITY).
 SQ SEQUENCE 585 AA: 65838 MW: F56DCA015F73451 CRG64;

Query Match 56.5%: Score 61: DB 1: Length 585;
 Best Local Similarity 77.8%: Pred. No. 3.25e-01;
 Matches 7: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

Db 321 TPEFLMSRK 329
 QY 1 TPEFLMSRN 9

RESULT 13
 ID ASN1 PEA STANDARD: PRT: 585 AA.
 AC P19251: 049925:
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1991 (Rel. 38, Last annotation update)
 DE ASPARAGINE SYNTHETASE, NODULE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4)
 DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).
 GN AS1.

OS Pisum sativum (Garden pea).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
 OC core eudicots: Rosidae: eurosids I: Fabales; Fabaceae; Papilionoideae;
 OC Pisum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. SPARKLE: TISSUE-ROOT NODULES;
 RX MEDLINE: 90151604.
 RA Tsai F.Y., Coruzzi G.M.;
 RT "Dark-induced and organ-specific expression of two asparagine
 RT synthetase genes in Pisum sativum.";
 RL EMBO J. 9:323-332(1990).
 RN [2]
 RP SEQUENCE OF 1-83 FROM N.A.
 RC STRAIN-CV. FELTHAM FIRST;
 RA Ngai N., Tsai F.Y., Coruzzi G.M.;
 RT "Light-induced transcriptional repression of the pea As1 gene:
 RT identification of cis-elements and transactors.";
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: to the EMBL/GenBank/DBJ databases.
 CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
 CC -1- TISSUE SPECIFICITY: NODULE.
 CC -1- INDUCTION: DARK-INDUCED.
 CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52179; CAA36429.1; -
 DR PIR: Y13321; CAA37362.1; -
 DR PIR: S11444; AJPANI.
 DR HSSP: P17169; IGMS.
 DR PROSITE: PS00443; GATASE_TYPE_II; 1.
 DR PFAM: PF00310; GATase_2; 1.
 DR PFAM: PF00733; Asn_synthase; 1.
 KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
 KM Multigene family.
 FT INIT_MET 0 BY SIMILARITY.
 FT ACT_SITE 1 GATASE (BY SIMILARITY).
 FT SITE 1 GATASE (BY SIMILARITY).
 SQ SEQUENCE 585 AA: 66222 MW: 6CEBC4417B120ECD CRG64;

Query Match 56.5%: Score 61: DB 1: Length 585;
 Best Local Similarity 77.8%: Pred. No. 3.25e-01;
 Matches 7: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

Db 322 TPEFLMSRK 330
 QY 1 TPEFLMSRN 9

RESULT 14
 ID ASNS MAIZE STANDARD: PRT: 585 AA.
 AC P49094:
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
 DE DEPENDENT ASPARAGINE SYNTHETASE).
 GN ASN1 OR AS.
 OS Zea mays (Maize).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC euphyllophytes: Spermatophyta: Magnoliophyta: Liliopsida; Poales;
 OC Poaceae; Zea.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-CV. DEA: TISSUE-ROOT MERISTEM;
 RA MEDLINE: 96158342.
 RX Chevalier C., Bourgeois E., Just D., Raymond P.;
 RT "Metabolic regulation of asparagine synthetase gene expression in
 RL maize (Zea mays L.) root tips.";
 CC Plant J. 9:1-11(1996).
 CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +
 CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
 CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
 CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X82849; CAA58052.1; -
 CC DR MAZEDB: 79071; -
 CC DR PROSITE: PS00443; GATASE_TYPE_II: 1.
 CC DR PFAM: PF00310; GATase_2: 1.
 CC DR PFAM: PF00733; Asn_synthase: 1.
 CC KW Ligase: Asparagine biosynthesis; Glutamine amidotransferase.
 CC FT INIT_MET 0 0 BY SIMILARITY.
 CC FT ACT_SITE 1 1 GATASE (BY SIMILARITY).
 CC SQ SEQUENCE 585 AA: 66446 MM: 095CE2F99973797E CRC64;
 CC
 CC Query Match 56.5%; Score 61; DB 1; Length 585;
 CC Best Local Similarity 77.8%; Pred. No. 3.25e-01;
 CC Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC Db 321 TPMFLMSRK 329
 CC QY 1 TPMFLMSRN 9
 CC
 CC RESULT 15
 CC ID ASNS_BRAOL STANDARD: PRT; 585 AA.
 CC AC P49091;
 CC DT 01-FEB-1996 (Rel. 33, Created)
 CC DT 01-FEB-1996 (Rel. 33, Last sequence update)
 CC DT 15-JUL-1998 (Rel. 36, Last annotation update)
 CC DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
 CC DEPENDENT ASPARAGINE SYNTHETASE).
 CC OS Brassica oleracea (Cauliflower).
 CC CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC CC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC CC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 CC CC Brassica.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RA STRAIN-CV. SHOGUN;
 CC RA Downs C.G., Pogson B.J., Davies K.M., Almira E.C.;
 CC RT "An asparagine synthetase cDNA clone from Broccoli (Brassica oleracea
 CC L.)".
 CC RL (in) Plant Gene Register PGR95-016.
 CC CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +
 CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
 CC CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
 CC CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X84448; CAA59138.1; -
 CC DR PROSITE: PS00443; GATASE_TYPE_II: 1.
 CC DR PFAM: PF00310; GATase_2: 1.
 CC DR PFAM: PF00733; Asn_synthase: 1.
 CC KW Ligase: Asparagine biosynthesis; Glutamine amidotransferase.
 CC FT INIT_MET 0 0 BY SIMILARITY.
 CC FT ACT_SITE 1 1 GATASE (BY SIMILARITY).
 CC SQ SEQUENCE 585 AA: 65541 MM: A73083898A22FE6 CRC64;
 CC
 CC Query Match 56.5%; Score 61; DB 1; Length 585;
 CC Best Local Similarity 77.8%; Pred. No. 3.25e-01;
 CC Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC Db 322 TPMFLMSRK 330
 CC QY 1 TPMFLMSRN 9
 CC
 CC Search completed: Sat May 13 07:12:43 2000
 CC Job time : 90 secs.

 W O R L D
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 07:13:02 2000; MasPar time 158.22 Seconds

Tabular output not generated. 6.573 Million cell updates/sec

Title: >US-08-991-628-4

Description: (1-15) from US08991628.pap

Sequence: 1 TPFMLSRNTGEVRT 15

Scoring table: PAM 150

Searched: 225878 segs, 69334122 residues

Post-processing: Minimum Match 0% Listing first 45 summaries

Database:

sptrembl12
 1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
 5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
 9:sp_phase 10:sp_plant 11:sp_protent 12:sp_unclassified
 13:sp_vertebrate 14:sp_virus

Statistics: Mean 27.384; Variance 31.843; scale 0.860

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|--------|------------------------|-----------|
| 1 | 90 | 83.3 | 993 | 11 | 035902 | DESMOGLEIN 3 (FRAGMENT | 4.79e-08 |
| 2 | 67 | 62.0 | 878 | 4 | 0915C2 | PROTODADHERIN GAMMA C5 | 2.77e-02 |
| 3 | 67 | 62.0 | 944 | 4 | 0915F6 | PROTODADHERIN GAMMA C5 | 2.77e-02 |
| 4 | 64 | 59.3 | 336 | 5 | 009661 | HYPOTHETICAL 37.7 KD P | 1.34e-01 |
| 5 | 63 | 58.3 | 291 | 2 | 004824 | ORF6 PROTEIN PRECURSOR | 2.24e-01 |
| 6 | 62 | 57.4 | 643 | 5 | 018290 | HYPOTHETICAL PROTEIN C | 3.72e-01 |
| 7 | 61 | 56.5 | 557 | 3 | 042902 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 8 | 61 | 56.5 | 578 | 10 | 0925T7 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 9 | 61 | 56.5 | 578 | 10 | 0925T6 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 10 | 61 | 56.5 | 581 | 10 | 0925T6 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 11 | 61 | 56.5 | 584 | 10 | 0925T3 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 12 | 61 | 56.5 | 585 | 10 | 065329 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 13 | 61 | 56.5 | 586 | 10 | 093618 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 14 | 61 | 56.5 | 586 | 10 | 040328 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 15 | 61 | 56.5 | 586 | 10 | 024483 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 16 | 61 | 56.5 | 871 | 4 | 0915C3 | PROTODADHERIN GAMMA C4 | 6.17e-01 |
| 17 | 61 | 56.5 | 938 | 4 | 0915F7 | PROTODADHERIN GAMMA C4 | 6.17e-01 |
| 18 | 59 | 54.6 | 148 | 1 | 058871 | 148AA LONG HYPOTHETICA | 1.66e+00 |
| 19 | 59 | 54.6 | 820 | 1 | 015039 | KIAA0327 PROTEIN. | 1.66e+00 |
| 20 | 59 | 54.6 | 832 | 1 | 091DLS | 852AA LONG HYPOTHETICA | 1.66e+00 |

| RESULT | ID | Score | Query Match | Length | DB | ID | Description | Pred. No. |
|--------|--------|-------|-------------|--------|----|--------|------------------------|-----------|
| 1 | 035902 | 90 | 83.3% | 993 | 11 | 035902 | DESMOGLEIN 3 (FRAGMENT | 4.79e-08 |
| 2 | 035902 | 67 | 62.0% | 878 | 4 | 0915C2 | PROTODADHERIN GAMMA C5 | 2.77e-02 |
| 3 | 035902 | 67 | 62.0% | 944 | 4 | 0915F6 | PROTODADHERIN GAMMA C5 | 2.77e-02 |
| 4 | 035902 | 64 | 59.3% | 336 | 5 | 009661 | HYPOTHETICAL 37.7 KD P | 1.34e-01 |
| 5 | 035902 | 63 | 58.3% | 291 | 2 | 004824 | ORF6 PROTEIN PRECURSOR | 2.24e-01 |
| 6 | 035902 | 62 | 57.4% | 643 | 5 | 018290 | HYPOTHETICAL PROTEIN C | 3.72e-01 |
| 7 | 035902 | 61 | 56.5% | 557 | 3 | 042902 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 8 | 035902 | 61 | 56.5% | 578 | 10 | 0925T7 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 9 | 035902 | 61 | 56.5% | 578 | 10 | 0925T6 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 10 | 035902 | 61 | 56.5% | 581 | 10 | 0925T6 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 11 | 035902 | 61 | 56.5% | 584 | 10 | 0925T3 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 12 | 035902 | 61 | 56.5% | 585 | 10 | 065329 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 13 | 035902 | 61 | 56.5% | 586 | 10 | 093618 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 14 | 035902 | 61 | 56.5% | 586 | 10 | 040328 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 15 | 035902 | 61 | 56.5% | 586 | 10 | 024483 | ASPARAGINE SYNTHETASE | 6.17e-01 |
| 16 | 035902 | 61 | 56.5% | 871 | 4 | 0915C3 | PROTODADHERIN GAMMA C4 | 6.17e-01 |
| 17 | 035902 | 61 | 56.5% | 938 | 4 | 0915F7 | PROTODADHERIN GAMMA C4 | 6.17e-01 |
| 18 | 035902 | 59 | 54.6% | 148 | 1 | 058871 | 148AA LONG HYPOTHETICA | 1.66e+00 |
| 19 | 035902 | 59 | 54.6% | 820 | 1 | 015039 | KIAA0327 PROTEIN. | 1.66e+00 |
| 20 | 035902 | 59 | 54.6% | 832 | 1 | 091DLS | 852AA LONG HYPOTHETICA | 1.66e+00 |

```

OC Eutheria: Primates: Catarrhini: Hominoidea: Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99308636.
RA WU Q., MANIATIS T.;
RT "A striking organization of a large family of human neural cadherin-
RT like cell adhesion genes."
RL Cell 97:779-790(1999).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AF152526; AAD43786.1; -.
DR PROSITE: PS00232; CADHERIN: 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ SEQUENCE 878 AA; 95192 MW; 07F807CA CRC32;

Query Match
Best Local Similarity 62.0%; Score 67; DB 4; Length 878;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 618 LFLVSTHTGEVRT 630
OY 3 MFLSRTNGEVRT 15

RESULT 3
ID 09Y5F6 PRELIMINARY; PRT; 944 AA.
AC 09Y5F6;
DT 01-NOV-1996 (TRENBLREL. 12, Created)
DT 01-NOV-1996 (TRENBLREL. 12, Last sequence update)
DE 01-NOV-1996 (TRENBLREL. 12, Last annotation update)
DE PROTOCOLADHERIN GAMMA C5.
GN PCDH-GAMMA-C5.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-BRAIN;
RX MEDLINE: 99308636.
RA WU Q., MANIATIS T.;
RT "A striking organization of a large family of human neural cadherin-
RT like cell adhesion genes."
RL Cell 97:779-790(1999).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AF152526; AAD43786.1; -.
DR PROSITE: PS00232; CADHERIN: 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ SEQUENCE 944 AA; 101920 MW; 2692C1D3 CRC32;

Query Match
Best Local Similarity 62.0%; Score 67; DB 4; Length 944;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 618 LFLVSTHTGEVRT 630
OY 3 MFLSRTNGEVRT 15

RESULT 4
ID 009661 PRELIMINARY; PRT; 336 AA.
AC 009661;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DE 01-MAY-1999 (TRENBLREL. 10, Last annotation update)
DE HYPOTHETICAL 37.7 KD PROTEIN ZK177.6 IN CHROMOSOME II.
GN ZK177.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RA ANDERSON K.;
SQ Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

```

```

CC -1- SIMILARITY: WEAK, TO YEAST PROTEIN CDC20.
DR EMBL: U21321; AAB36970.1; -.
DR WORMPEP: ZK177.6; CE02095.
DR PFM: PFO0400; WD40; 1.
KW Hypothetical protein.
SQ SEQUENCE 336 AA; 37668 MW; 2DC77B26 CRC32;

Query Match
Best Local Similarity 59.3%; Score 64; DB 5; Length 336;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 187 PMFTDEHTGCVRT 200
OY 2 PMFLSRTNGEVRT 15

RESULT 5
ID 004824 PRELIMINARY; PRT; 291 AA.
AC 004824;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DE 01-JAN-1999 (TRENBLREL. 09, Last annotation update)
DE ORF6 PROTEIN PRECURSOR.
OS Salmonella typhimurium.
OC plasmid 90 kb virulence.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-LT2;
RX MEDLINE: 93316852.
RA FRIEDRICH M.J., KINSEY N.E., VILA J., KADNER R.J.;
RT "Nucleotide sequence of a 13.9 kb segment of the 90 kb virulence
RT plasmid of Salmonella typhimurium: the presence of fimbrial
RT biosynthetic genes."
RL MOL. Microbiol. 8:543-558(1993).
DR EMBL: L08613; AAC36963.1; -.
KW plasmid; Signal.
FT SIGNAL 1 23
FT CHAIN 24 291 ORF6 PROTEIN.
SQ SEQUENCE 291 AA; 30683 MW; CAD5ADDF CRC32;

Query Match
Best Local Similarity 58.3%; Score 63; DB 2; Length 291;
Matches 12; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 128 TLMFSLR-TGEVRT 141
OY 1 TPEFLSRTNGEVRT 15

RESULT 6
ID 018290 PRELIMINARY; PRT; 643 AA.
AC 018290; Q20732;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DE 01-MAY-1999 (TRENBLREL. 10, Last annotation update)
DE HYPOTHETICAL PROTEIN C29A12.6.
GN C29A12.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA WILKINSON J.;
SQ Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z73970; CAA98248.1; -.
DR EMBL: Z77664; CAA98248.1; JOINED.
DR EMBL: Z77664; CAB01217.1; -.
DR EMBL: Z73970; CAB01217.1; JOINED.
SQ SEQUENCE 643 AA; 73571 MW; C0092D81 CRC32;

Query Match
Best Local Similarity 57.4%; Score 62; DB 5; Length 643;

```

Best Local Similarity 53.8% Pred. No. 3,72e-01;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 532 MFLMRHGTTRT 544
|||:|:|:|:
OY 3 MFLSRNTGEVRT 15

RESULT 7
ID 042902; PRELIMINARY; PRT: 557 AA.

AC 042902;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE ASPARAGINE SYNTHETASE.
GN SPBC119.10.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiaascmycetaceae;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;
RA WOOD V., RAJANDREAM M.A., BARRELL B.G., SKELTON J., CHURCHER C.M.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF022117; CAAL7925.1; -;
DR PFAM: PF00733; Asn-synthase; 1.
DR PFAM: PF00310; GATase_2; 1.
SQ SEQUENCE 557 AA; 63240 MW; 1805CB15 CRC32;

Query Match 56.5% Score 61; DB 3; Length 557;
Best Local Similarity 77.8% Pred. No. 6,17e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 335 TPEYLLSRK 343
|||:|:|:|:
OY 1 TPEYLLSRN 9

RESULT 8
ID 092577; PRELIMINARY; PRT: 578 AA.

AC 092577;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE ASPARAGINE SYNTHETASE.

GN ASN3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsids.

RN [1]
RP SEQUENCE FROM N.A.
RA LAM H.M., HSIEH M.H., CORUZZI G.M.;
RT "Reciprocal regulation of distinct asparagine synthetase genes by
RT light and metabolites in Arabidopsis thaliana.";
RL Plant J. 0:0-0(1998).

DR EMBL: AF095452; AAC72836.1; -;
DR MENDEL: 34482; Arath.1042;34482.
KM PFAM: PF00310; GATase_2; 1.
SQ SEQUENCE 578 AA; 65223 MW; 8C46069E CRC32;

Query Match 56.5% Score 61; DB 10; Length 578;
Best Local Similarity 77.8% Pred. No. 6,17e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 322 TPEYLLSRK 330
|||:|:|:|:
OY 1 TPEYLLSRN 9

RESULT 9
ID 092576; PRELIMINARY; PRT: 578 AA.

AC 092576;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE ASPARAGINE SYNTHETASE.

GN ASN2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsids.

RN [1]
RP SEQUENCE FROM N.A.
RA LAM H.M., HSIEH M.H., CORUZZI G.M.;
RT "Reciprocal regulation of distinct asparagine synthetase genes by
RT light and metabolites in Arabidopsis thaliana.";
RL Plant J. 0:0-0(1998).
DR EMBL: AF095453; AAC72837.1; -;
DR MENDEL: 34483; Arath.1042;34483.
SQ SEQUENCE 578 AA; 65077 MW; DD3E9594 CRC32;

Query Match 56.5% Score 61; DB 10; Length 578;
Best Local Similarity 77.8% Pred. No. 6,17e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 322 TPEYLLSRK 330
|||:|:|:|:
OY 1 TPEYLLSRN 9

RESULT 10
ID P93167; PRELIMINARY; PRT: 581 AA.

AC P93167;
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE ASPARAGINE SYNTHETASE 2 (EC 6.3.5.4) (ASPARAGINE SYNTHASE
DE (GLUTAMINE-HYDROLYSING)) (ASPARAGINE SYNTHETASE
DE (GLUTAMINE-HYDROLYSING)).

GN AS2.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Glycine.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CENTURY; TISSUE-MATURE LEAF;
RX MEDLINE: 97188563.
RA HUGHES C.A., BEARD H., MATTHEWS B.F.;
RT "Molecular cloning and expression of two cDNAs encoding asparagine
RT synthetase in soybean.";
RL Plant Mol. Biol. 33:301-311(1997).
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC DIPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.

DR EMBL: U77678; AAC49613.1; -;
DR HSSP: P17169; IGMS.
DR MENDEL: 8141; Glyma.1042;8141.
DR PFAM: PF00733; Asn-synthase; 1.
DR PFAM: PF00310; GATase_2; 1.
KM LIGase.

Query Match 56.5% Score 61; DB 10; Length 581;
Best Local Similarity 77.8% Pred. No. 6,17e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 322 TPEYLLSRK 330
|||:|:~:~:~:
OY 1 TPEYLLSRN 9

RESULT 11

ID 092523 PRELIMINARY; PRT: 584 AA.
 AC 092523;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE ASPARAGINE SYNTHETASE TYPE II (EC 6.3.5.4).
 GN AS2.
 OS Phaseolus vulgaris (Kidney bean) (French bean).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
 CC Phaseolus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. GREAT NORTHERN; TISSUE-ROOT;
 RA OSUNA D., GALVEZ G., PINEDA M., AGUILAR M.;
 RT "R-PCR cloning and expression of a cDNA encoding a type-II asparagine
 synthetase."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ009952; CA08913.1; -
 DR HSSP: P17169; IGMS.
 DR MENDEL: 34839; Pnavu:1042;34839.
 KW Ligase.
 SQ SEQUENCE 584 AA; 65743 MW; CDE20571 CRC32;

Query Match 56.5%; Score 61; DB 10; Length 584;
 Best Local Similarity 77.8%; Pred. No. 6.17e-01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPFMLSRK 330
 QY 1 TPFMLSRN 9

RESULT 12
 ID 065329 PRELIMINARY; PRT: 585 AA.
 AC 065329;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE ASPARAGINE SYNTHETASE.
 GN AS.
 OS Eleagnus umbellata.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC core eudicots; Rosidae; eurosids I; Rosales; Elaeagnaceae; Elaeagnus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ROOT NODULE;
 RA KIM H.-B., AN C.-S.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF061740; AAC16325.1; -
 DR MENDEL: 28685; elaum:1042;29685.
 DR PFAM: PF00733; Asn_synthase; 1.
 DR PFAM: PF00310; GATase_2; 1.
 SQ SEQUENCE 585 AA; 65878 MW; 8A12777D CRC32;

Query Match 56.5%; Score 61; DB 10; Length 585;
 Best Local Similarity 77.8%; Pred. No. 6.17e-01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPFMLSRK 330
 QY 1 TPFMLSRN 9

RESULT 13
 ID P93618 PRELIMINARY; PRT: 586 AA.
 AC P93618;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE ASPARAGINE SYNTHETASE (EC 6.3.5.4) (ASPARAGINE SYNTHASE

DE (GLUTAMINE-HYDROLYSING)) (ASPARAGINE SYNTHETASE
 DE (GLUTAMINE-HYDROLYSING)).
 GN VFAS1.
 OS Vicia faba (broad bean).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
 CC Vicia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KLEINE THUNDERINGER; TISSUE-ROOT NODULE;
 RA KUSTER H., ALBUS U., FRUEHLING M., TCHETKOVA S.A., TIKHONOVITCH I.A.,
 RA PUEHLER A., PERLICK A.M.;
 RL Plant Sci. 124:89-95(1997).
 CC -I- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +
 CC DIPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
 DR EMBL: Z72354; CA96526.1; -
 DR HSSP: P17169; IGMS.
 DR MENDEL: 11023; Vlcfa:1042;11023.
 DR PFAM: PF00733; Asn_synthase; 1.
 DR PFAM: PF00310; GATase_2; 1.
 KW Ligase.
 SQ SEQUENCE 586 AA; 66236 MW; AD8BD927 CRC32;

Query Match 56.5%; Score 61; DB 10; Length 586;
 Best Local Similarity 77.8%; Pred. No. 6.17e-01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 323 TPFMLSRK 331
 QY 1 TPFMLSRN 9

RESULT 14
 ID 040328 PRELIMINARY; PRT: 586 AA.
 AC 040328;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE ASPARAGINE SYNTHETASE.
 GN Medicago sativa (Alfalfa).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
 CC Medicago.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. SARANAC;
 RA GANTT S.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: LA0327; ABA46058.1; -
 DR HSSP: P17169; IGMS.
 DR MENDEL: 9036; Medsa:1042;9036.
 DR PFAM: PF00733; Asn_synthase; 1.
 DR PFAM: PF00310; GATase_2; 1.
 SQ SEQUENCE 586 AA; 66462 MW; 962C3A3F CRC32;

Query Match 56.5%; Score 61; DB 10; Length 586;
 Best Local Similarity 77.8%; Pred. No. 6.17e-01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 323 TPFMLSRK 331
 QY 1 TPFMLSRN 9

RESULT 15
 ID 024483 PRELIMINARY; PRT: 586 AA.
 AC 024483;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE ASPARAGINE SYNTHETASE.

OS Medicago sativa (Alfalfa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Medicago.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97432147.
RA SHI L., TWARY S.N., YOSHIOKA H., GREGGSON R.G., MILLER S.S.,
RA SAMAC D.A., GANTT J.S., UNKEFER P.J., VANCE C.P.;
RT "Nitrogen assimilation in alfalfa: isolation and characterization of
RT an asparagine synthetase gene showing enhanced expression in root
RT nodules and dark-adapted leaves.";
RL Plant Cell 9:1339-1356(1997).
DR EMBL: U89923; AAB81011.1; -.
DR HSSP: P17169; IGMS.
DR MENDEL: 27006; Medsa:1042;27006.
DR PFAM: PF00733; Asn-synthase; 1.
DR PFAM: PF00310; GATase-2; 1.
SQ SEQUENCE 586 AA; 66448 MW; 63C8B6BA CRC32;

Query Match 56.58; Score 61; DB 10; Length 586;
Best Local Similarity 77.88; Pred. No. 6.17e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 323 TPWFLMSRK 331
QY 1 TPWFLLSRN 9

Search completed: Sat May 13 07:15:54 2000
Job time : 172 secs.

THIS PAGE BLANK (USPTO)

Mon May 15 08:08:55 2000

RESULT 2
 R30742 standard: Protein: 999 AA.
 R30742: 14-JUN-1993 (first entry)
 Human pemphigus vulgaris 130kd antigen.
 E pemphigus vulgaris: skin disease; autoantibodies:
 W keratinocyte cell surface antigen; glycoprotein; cell adhesion.
 N Homo sapiens.
 US7798918-A.
 15-DEC-1992.
 27-NOV-1991: 798918.
 (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PA Amagel M, Klaus-Kovtun V, Stanley JR:
 WPI: 93-067436/08.
 DR N-PSDB: Q35992.
 PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
 PT diagnostic and therapeutic uses
 PS disclosure: Fig 7: 50pp; English.
 CC This sequence is the pemphigus vulgaris 130kd antigen. The protein
 CC and its encoding DNA may be used in the diagnosis and treatment of
 CC pemphigus vulgaris. It is thought that the antigen may be a cell
 CC adhesion molecule.
 SQ Sequence 999 AA:
 Query Match 100.0%: Score 113; DB 1; Length 999;
 Best Local Similarity 100.0%: Pred. No. 3.35e-05; Indels 0; Gaps 0;
 Matches 15; Conservative
 Db 251 CECNIKVDVNDNFP 265
 1 CECNIKVDVNDNFP 15
 |||||
 QY 1 CECNIKVDVNDNFP 15

RESULT 3
 W07908 standard: Protein: 614 AA.
 AC W07908:
 29-JAN-1997 (first entry)
 Pemphigus vulgaris antigen: protein extracellular region;
 DE Autoantibody: immunoglobulin G; IgG1; fusion protein; diagnosis;
 KW treatment: pemphigus vulgaris; PV; bulla; blister; skin disease;
 KM dermatology.
 OS Homo sapiens.
 PN J08188540-A.
 23-JUL-1996.
 30-JUN-1995: 165632.
 30-JUN-1994: JP-173291.
 PR (NISH/) NISHIKAWA T.
 PA WPI: 96-388562/39.
 DR Fused protein recognised by pemphigus vulgaris autoantibody -
 PT useful to treat and diagnose pemphigus vulgaris
 PS Claim 1: Page 7-9: 9pp; Japanese.
 CC W07908 represents the human pemphigus vulgaris (PV) antigen
 CC extracellular region. The PV antigen is produced in patients with
 CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
 CC relapsing disease causing suppurative, intra-epidermal bullae
 CC (vesicles) of the skin and mucous membranes, which is fatal if
 CC untreated. The PV antigen was fused to a human IgG1 hinge region
 CC and the resulting fusion protein is useful to treat or diagnose
 CC pemphigus vulgaris.
 SQ Sequence 614 AA:
 Query Match 90.3%: Score 102; DB 1; Length 614;
 Best Local Similarity 93.3%: Pred. No. 6.39e-04; Indels 0; Gaps 0;
 Matches 14; Conservative
 Db 250 CECNIKVDVNDNFP 264
 1 CECNIKVDVNDNFP 15
 |||||
 QY 1 CECNIKVDVNDNFP 15

ID W15489 standard: Protein: 778 AA.
 AC W15489:
 17-JUN-1997 (first entry)
 Pemphigus foliaceus antigen-IgG constant region fusion protein.
 DE Pemphigus foliaceus: autoantibody; constant region; IgG;
 KW extracellular region; antigen; hinge portion; skin;
 KM dermatitis herpetiformis; fusion protein; detection; ss.
 OS Homo sapiens.
 CH Chimeric - Homo sapiens.
 FT Key Location/Qualifiers
 1..345
 /note="pemphigus foliaceus antigen protein"
 FT domain
 FT J09077800-A.
 PA 25-MAR-1997.
 PD 12-SEP-1995: 260899.
 PR 12-SEP-1995: JP-260899.
 PA (NISH/) NISHIKAWA T.
 DR WPI: 97-241758/22.
 DR P-PSDB: T66428.
 PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
 PT through the hinge region used to treat pemphigus foliaceus
 PS Claim 1: Page 10-12: 17pp; Japanese.
 CC This sequence represents a fused protein recognised by pemphigus
 CC foliaceus patient autoantibody which comprises the constant region
 CC of IgG linked to the extracellular region of pemphigus foliaceus
 CC antigen protein through the hinge portion. Pemphigus foliaceus
 CC a chronic, generalised, vesicular and scaling skin eruption fusion
 CC to dermatitis herpetiformis. The pemphigus foliaceus antigen is
 CC protein is useful to treat pemphigus foliaceus. The antigen is
 CC especially administered via a carrier. The fusion protein is also
 CC useful for detecting pemphigus foliaceus antibodies which is useful
 CC in immunodiagnosis. The fusion protein has little or no side effects.
 SQ Sequence 778 AA:
 Query Match 78.8%: Score 89; DB 1; Length 778;
 Best Local Similarity 80.0%: Pred. No. 1.95e-02; Indels 0; Gaps 0;
 Matches 12; Conservative
 Db 253 CECNIKVDVNDNFP 267
 1 CECNIKVDVNDNFP 15
 |||||
 QY 1 CECNIKVDVNDNFP 15

RESULT 5
 R86865 standard: Protein: 787 AA.
 AC R86865:
 27-AUG-1996 (first entry)
 Human protocadherin pc3.
 DE Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
 KW catenin; therapy.
 OS Homo sapiens.
 PN M09600289-A1.
 PD 04-JAN-1996.
 PF 26-JUN-1995: U08071.
 27-JUN-1994: US-268161.
 PA (DOHE-) DOHENY EYE INST.
 PI SUZUKI S.
 DR WPI: 96-068873/07.
 DR N-PSDB: T03572.
 PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
 PT pc5 - involved in cell-cell adhesion and regulation activities
 PS Claim 1: Page 115-119: 146pp; English.
 CC R86865 represents the human protocadherin pc3. These sequences are
 CC related to cadherin, and possess cell adhesive ability. Cadherins are
 CC glycosylated integral membrane proteins that are involved in cell-cell
 CC adhesion. Cadherins are composed of an N-terminal extracellular domain
 CC which consists of 5 unique subdomains, a membrane spanning domain, and a
 CC C-terminal cytoplasmic domain. The cytoplasmic domain interacts with the
 CC cytoskeleton through catenins and other cytoskeleton associated proteins.
 CC The cytoplasmic domain is not present in all cadherins, but in those
 CC which possess it, it is essential for the cadherins adhesive function.
 CC The cadherins which do not possess a cytoplasmic domain appear to

CC proteins also allows investigation of the structure and function of
CC cadherin. The cadherin proteins may be isolated by using anti-cadherin
CC antibodies. These antibodies may also be used to modulate the activity
CC of cadherin and to determine the tissue specific distribution of cadherin
CC proteins. Each subclass of cadherins has a unique tissue distribution
CC pattern.
SQ Sequence 780 AA;

Query Match 50.4%; Score 57; DB 1; Length 780;
Best Local Similarity 81.8%; Pred. No. 5.45e+01;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 135 IKVNDVNDMP 145
IIIIIIII
OY 5 IKVNDVNDMP 15

RESULT 9
ID W84311 standard; Protein: 83 AA.

AC W84311;
DT 18-MAR-1999 (first entry)
DE Oryza sativa sequence used to exemplify the invention.
KW DNA sequence comparison.
OS Oryza sativa.
PN J10334104-A.
PD 18-DEC-1998.
PF 05-MAR-1998; 053102.
PR 31-MAR-1997; JP-079586.
PA (HITA) HITACHI LTD.
DR WPI: 99-110965/10.

PT Comparison of DNA base sequences - using accurate observations of
PT insertions and deletions

PS Disclosure: Page 16-17; 24pp; Japanese.

CC The present sequence is used to demonstrate the method of the invention.
CC The specification describes the comparison of DNA base sequences. The
CC method involves 3 steps of: (1) division of a first and second nucleic
CC acid sequence into base groups of three base lengths, and translation
CC into amino acids; (2) all base and amino acid insertions and deletions
CC of the two nucleic acid sequences are observed and the amino acid
CC sequences encoded by these nucleic acid sequences are compared; and
CC (3) adjacent bases and amino acids are compared and step (1) repeated
CC after shifting the translation frame by one base towards the 3' end or
CC shifting the translation frame one base towards the 3' end, but
CC encompassing four bases, and not translating the second or third base.
CC The new method is useful for comparing DNA base sequences.
SQ Sequence 83 AA;

Query Match 49.6%; Score 56; DB 1; Length 83;
Best Local Similarity 40.0%; Pred. No. 6.88e+01;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 68 CESSIYOYIDEARF 82
IIIIIIII
OY 1 CECNIKVKVDNDMP 15

RESULT 10
ID R55273 standard; Protein: 1822 AA.

AC R55273;
DT 31-JAN-1995 (first entry)
DE Beta subunit of integrin cell surface receptor.
KW Integrin; alpha; beta; subunit; glycoprotein; heterodimer;
KW transmembrane; extracellular matrix; cell signalling; cytoskeleton;
KW behaviour; signal transduction; receptor.
OS Homo sapiens.

FN Key Location/Qualifiers
FT peptide 28..42

FT modified_site 491 /note- "N-terminal peptide."

FT modified_site 617 /note- "Potential N-linked glycosylation site."

FT modified_site 695 /note- "Potential N-linked glycosylation site."

FT domain /note- "Potential N-linked glycosylation site."
FT 711..733
FT modified_site 980 /note- "Putative transmembrane domain."
FT modified_site 1593 /note- "Potential N-linked glycosylation site."
FT modified_site 1593 /note- "Potential N-linked glycosylation site."

PN US5320942-A.
PD 14-JUN-1994.
PF 19-FEB-1987; 016552.
PR 19-FEB-1987; US-016552.
PR 04-JAN-1989; US-293384.
PR 01-OCT-1990; US-591105.

PA (KAIJ/) KAJIUI S.
PA (QVAR/) QVARANTA V.
PI Kaiji S, Quaranta V;
DR WPI: 94-191533/23.
DR N-PSDB: 065674.

PT Diagnosing presence of abnormal epithelial tissue in vitro -
PT utilises monoclonal antibodies to alpha6 beta4 cell surface protein

PS Example 5; Figure 9; 34pp; English.

CC Integrins are heterodimers comprised of alpha and beta subunits, that
CC are non-covalently associated transmembrane glycoproteins. 11 alpha
CC chains and 6 beta chains have been recognised in man. Each alpha
CC subunit tends to associate with only one type of beta subunit but
CC there are several exceptions. Integrins mediate (in part) the
CC interaction of cells with the extracellular matrix, forming a link
CC between the extracellular matrix and the cytoskeleton. They may
CC transmit signals from the extracellular to the intracellular
CC environment, affecting cell behaviour. This sequence is the beta4
CC subunit of an alpha6 beta4 integrin.
SQ Sequence 1822 AA;

Query Match 49.6%; Score 56; DB 1; Length 1822;
Best Local Similarity 63.6%; Pred. No. 6.88e+01;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 650 ECFNKVKVND 660
IIIIIIII
OY 2 ECFNKVKVND 12

RESULT 11
ID R87107 standard; Peptide: 43 AA.

AC R87107;
DT 28-AUG-1996 (first entry)
DE Protocadherin clone RAT-224.
KW Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
KW catenin; therapy; clone.
OS Rattus rattus.
PN W09600288-A1.
PD 04-JAN-1996.
PF 26-JUN-1995; U08071.
PR 27-JUN-1994; US-268161.
PA (DOHE-) DOHENY EYE INST.

PI Suzuki S;
DR WPI: 96-068873/07.
DR N-PSDB: T03582.

PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
PT pc5 - involved in cell-cell adhesion and regulation activities

PS Example 1; Page 40; 146pp; English.

CC R87107-R87109 represent partial fragments of the rat protocadherin
CC sequence. The cDNAs encoding these sequences were isolated after
CC screening a rat brain cDNA preparation with the primers shown in T03575
CC and T03576. The primers were constructed from portions of the amino acid
CC sequences of the third and fourth extracellular domains of published
CC cadherin sequences. The full length cDNA sequence encoding rat
CC protocadherin pc5 is represented in T03574. The cytoplasmic domain of
CC cadherin interacts with the cytoskeleton through catenins and other
CC cytoskeleton associated proteins. The cytoplasmic domain is not present
CC in all cadherins, but in those which possess it, it is essential for the
CC cadherin adhesive function. The cadherins which do not possess a
CC cytoplasmic domain appear to function via a different method from those

CC with a cytoplasmic domain. These protein sequences are involved in
 CC cell-cell adhesion. These sequences may have regulatory functions in the
 CC cell, as well as the cell-cell adhesive properties. Antibodies produced
 CC against these sequences are useful for modulating the binding activity of
 CC these procadherins, and can be used therapeutically.
 SO Sequence 43 AA:

Query Match 47.8% Score 54: DB 1: Length 43:
 Best Local Similarity 66.7%: Pred. No. 1.09e+02:
 Matches 10: Conservative 0: Mismatches 5: Indels 0: Gaps 0:

Db 29 CKIKVVDVNDNAP 43
 | ||| ||||| |
 Qy 1 CECNKKVVDVNDNFP 15

RESULT 12
 ID R58865 standard; Protein: 43 AA.
 AC R58865:
 DT 17-APR-1995 (first entry)
 DE Rac-224 cadherin partial sequence.
 KW Cadherin: cell adhesion molecule.
 PN MO9414960-A.
 PD 07-JUL-1994.
 PR 23-DEC-1993: U12588.
 PR 29-DEC-1992: US-998003.
 PA (DOHE-) DOHEM EYE INST.
 PI Suzuki S:
 DR WPI: 94-293849/36.
 DR N-PSDB: 068957.
 PT Polynucleotide sequences encoding new proto:cadherins - useful
 PT for modulating natural binding and regulating activities.
 PS Example: Page 38, 114pp, English.
 CC Two regions of conserved AA sequence, one from the middle of the
 CC third cadherin extracellular subdomain (EC-3) and the other from the
 CC C-terminus of the fourth extracellular subdomain (EC-4) were
 CC identified. The corresp. degenerate oligos (Q68949, Q68950) were
 CC designed for use as PCR primers. PCR was carried out on a rat brain
 CC cDNA prep. Two major bands of about 450 bps and 130 bps were found.
 CC The 450 bp band corresponded to the expected length between the two
 CC primer sites, but the 130 bp band could not be predicted from any
 CC of the previously identified cadherin sequences. The 450 bp and 130
 CC bp bands were extracted and sequenced. Nineteen novel partial cDNA
 CC clones were isolated. The DNA and deduced AA sequences of the
 CC clones (including sequences corresp. to the PCR primers) are given
 CC in Q68951-Q68969 and R58860-R58878. The deduced AA sequences of the
 CC cDNA clones are homologous to, but distinct from the known
 CC cadherins. The cadherins described thus far have highly conserved
 CC short AA sequences in the EC-3 including the consensus sequence
 CC D-Y-E or D-E-F located at the middle region of the subdomain and
 CC the consensus sequence in R58879 or R58880 at its end, while the
 CC corresp. sequences of other subdomains, except for the 5th extra-
 CC cellular subdomain (EC-5), are D-R-E and the sequence in R58881
 CC respectively. In contrast the deduced AA sequences of the new
 CC clones that corresp. to cadherin extracellular subdomains include
 CC the sequence D-Y-E or D-F-E at one end, but have the sequence
 CC D-X-N-D-N-X-P-X-F instead of R58879 or R58880 at the other end.
 CC The polypeptides encoded by the partial clones are homologous to
 CC previously identified cadherins but did not show significant
 CC homology to any other sequences in Genbank. Therefore, the partial
 CC cDNAs appear to comprise a new subclass of cadherin-related
 CC molecules.
 SO Sequence 43 AA:

Query Match 47.8% Score 54: DB 1: Length 43:
 Best Local Similarity 66.7%: Pred. No. 1.09e+02:
 Matches 10: Conservative 0: Mismatches 5: Indels 0: Gaps 0:

Db 29 CKIKVVDVNDNAP 43
 | ||| ||||| |
 Qy 1 CECNKKVVDVNDNFP 15

RESULT 13
 ID W55470 standard; Protein: 83 AA.
 AC W55470:
 DT 24-JUN-1998 (first entry)
 DE H. pylori ORF 06ap1119.16594193.fl.9 secreted protein.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW Identification; binding compound; bacteria; life cycle; activator;
 KW Inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 OS Helicobacter pylori.
 PN MO9737044-A1.
 PD 09-OCT-1997.
 PR 27-MAR-1997: U05223.
 PR 06-DEC-1996: US-761218.
 PR 29-MAR-1996: US-625811.
 PR 02-APR-1996: US-758731.
 PR 25-OCT-1996: US-736905.
 PR 28-OCT-1996: US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D:
 DR WPI: 97-503122/46.
 DR N-PSDB: V24879.
 PT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claims 14, 94; Page 677-678, 1145pp, English.

CC This sequence is a H. pylori secreted protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds. The
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 SO Sequence 83 AA:

Query Match 46.9% Score 53: DB 1: Length 83:
 Best Local Similarity 50.0%: Pred. No. 1.37e+02:
 Matches 5: Conservative 3: Mismatches 2: Indels 0: Gaps 0:

Db 43 CENCNKKIPSI 52
 | : ||||| :
 Qy 1 CECNKKVVDVNDNFP 10

RESULT 14
 ID W55280 standard; Protein: 87 AA.
 AC W55280:
 DT 02-JUL-1998 (first entry)
 DE H. pylori ORF 11ge10309orf7 protein.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW Identification; binding compound; bacteria; life cycle; activator;
 KW Inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 OS Helicobacter pylori.
 PN MO9737044-A1.
 PD 09-OCT-1997.
 PR 27-MAR-1997: U05223.
 PR 06-DEC-1996: US-761218.
 PR 29-MAR-1996: US-625811.
 PR 02-APR-1996: US-758731.
 PR 25-OCT-1996: US-736905.

PR 28-OCT-1996: US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR WPI: 97-503122/46.
 DR N-PSDB: V24689.
 PT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claim 14: Page 516: 1145pp: English.
 CC This sequence is a H. pylori protein of unspecified function.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds.
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 SQ Sequence 87 AA;

OY : : 1:|||||
 5 IKVKDVNDNFP 15

Search completed: Sat May 13 07:22:09 2000
 Job time : 7 secs.

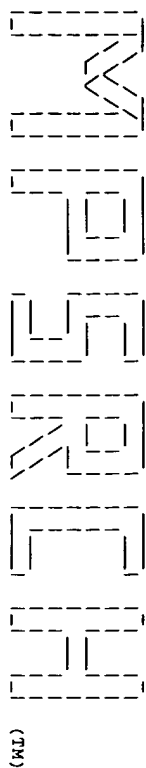
Query Match 46.9%; Score 53; DB 1; Length 87;
 Best Local Similarity 50.0%; Pred. No. 1.37e+02;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 47 CNCNIKIPST 56
 1:||||| :
 OY 1 CECNIKVKDV 10

RESULT 15
 ID W13139 standard; Protein: 11 AA.
 AC W13139;
 DT 14-MAY-1997 (first entry)
 DE Human cadherin-5 antigenic epitope (residues 242-252).
 KW Ca2+ dependent; cell adhesion protein; cadherin; antibody;
 KW purification; determination; epitope; tissue expression;
 KW binding antagonist; calcium ion; antigen.
 OS Homo sapiens.
 PN US5597725-A.
 PD 28-JAN-1997.
 PE 17-APR-1992: 872643.
 PR 17-APR-1992: US-872643.
 PR 19-APR-1993: US-049460.
 PR 26-JAN-1994: US-188228.
 PA (DOHE-) DOHEMY EYE INST.
 PI Suzuki S;
 DR WPI: 97-108328/10.
 PT Antibodies to cadherin proteins - useful as cadherin antagonists,
 PT etc.
 PS Claim 5; Column 112; 59pp: English.
 CC The present sequence is an antigenic epitope from human cadherin-5,
 CC which is a Ca2+ dependent cell adhesion protein. Antibodies or
 CC fragments that specifically bind the epitope can be used to purify
 CC the cadherin, determine its tissue expression and antagonise its
 CC ligand/antiligand binding activities.
 SQ Sequence 11 AA;

Query Match 46.0%; Score 52; DB 1; Length 11;
 Best Local Similarity 54.5%; Pred. No. 1.72e+02;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 1 VTLDINDNFP 11



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:21:36 2000; MasPar time 4.30 Seconds
164.428 Million cell updates/sec
Tabular output not generated.

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pep
Perfect Score: 113
Sequence: 1 CECNKKVDVNDNFP 15

Scoring table:
PAM 150
Gap 15

Searched: 142080 segs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 25.922; Variance 35.935; scale 0.721

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----------|--------------------------|-----------|
| 1 | 113 | 100.0 | 999 | 1 IJHUG3 | desmoglein 3 precursor | 1.81e-12 |
| 2 | 89 | 78.8 | 1043 | 1 IJBOG1 | desmoglein 1 precursor | 9.39e-07 |
| 3 | 89 | 78.8 | 1049 | 1 IJHUG1 | desmoglein 1 precursor | 9.39e-07 |
| 4 | 66 | 58.4 | 171 | 2 S62484 | hypothetical protein | 8.66e-02 |
| 5 | 63 | 55.8 | 394 | 2 H64448 | polyferredoxin - Meth | 3.40e-01 |
| 6 | 60 | 53.1 | 809 | 1 IJBODD | desmocollin 2b precursor | 1.28e+00 |
| 7 | 60 | 53.1 | 863 | 1 IJBODC | desmocollin 2a precursor | 1.28e+00 |
| 8 | 60 | 53.1 | 866 | 2 I45858 | desmocollin - bovine | 1.28e+00 |
| 9 | 59 | 52.2 | 709 | 2 T16176 | hypothetical protein | 1.98e+00 |
| 10 | 59 | 52.2 | 2163 | 2 T15276 | hypothetical protein | 3.05e+00 |
| 11 | 58 | 51.3 | 527 | 2 S62484 | hypothetical protein | 3.05e+00 |
| 12 | 58 | 51.3 | 832 | 2 S55396 | LI-cadherin - human | 3.05e+00 |
| 13 | 58 | 51.3 | 1324 | 2 S51622 | cut3 protein - fission | 3.05e+00 |
| 14 | 57 | 50.4 | 235 | 2 E70378 | DNA replication prote | 4.66e+00 |
| 15 | 57 | 50.4 | 666 | 2 T10664 | serine/threonine-spec | 4.66e+00 |
| 16 | 57 | 50.4 | 682 | 2 T06106 | proliferating-cell nu | 4.66e+00 |
| 17 | 57 | 50.4 | 794 | 1 IJHUC5 | cadherin 5 precursor | 4.66e+00 |
| 18 | 57 | 50.4 | 829 | 1 I46536 | KSP-cadherin - rabbit | 4.66e+00 |
| 19 | 57 | 50.4 | 847 | 1 IJHUBD | desmocollin 3a precursor | 4.66e+00 |
| 20 | 57 | 50.4 | 901 | 1 IJHUBA | desmocollin 3b precursor | 4.66e+00 |
| 21 | 57 | 50.4 | 1979 | 2 C71622 | hypothetical protein | 4.66e+00 |
| 22 | 56 | 49.6 | 413 | 1 DEACNT | D-nopaline dehydriogen | 7.09e+00 |
| 23 | 56 | 49.6 | 787 | 2 S68699 | potassium channel pro | 7.09e+00 |

| | | | | | |
|----|------|------|----------|------------------------|----------|
| 24 | 49.6 | 964 | 2 JC5545 | integrin betaE - hum | 7.09e+00 |
| 25 | 49.6 | 1807 | 2 JC6319 | integrin beta-4 chain | 7.09e+00 |
| 26 | 49.6 | 1875 | 2 A36429 | integrin beta-4 chain | 7.09e+00 |
| 27 | 48.7 | 296 | 2 A71232 | hypothetical protein | 1.07e+01 |
| 28 | 48.7 | 376 | 2 E70361 | chaperone DnaJ - Aqu | 1.07e+01 |
| 29 | 48.7 | 496 | 2 H75122 | sarcosine oxidase, ch | 1.07e+01 |
| 30 | 48.7 | 902 | 2 T00568 | hypothetical protein | 1.07e+01 |
| 31 | 48.7 | 1389 | 2 I50090 | carboxypeptidase gp18 | 1.07e+01 |
| 32 | 47.8 | 108 | 2 C72231 | ATP synthase F1, subu | 1.62e+01 |
| 33 | 47.8 | 249 | 2 G64415 | hypothetical protein | 1.62e+01 |
| 34 | 47.8 | 493 | 2 E71008 | hypothetical protein | 1.62e+01 |
| 35 | 47.8 | 671 | 2 A45730 | phenol 2-monooxygenas | 1.62e+01 |
| 36 | 47.8 | 807 | 2 T12177 | potassium channel pro | 1.62e+01 |
| 37 | 47.8 | 1790 | 2 S67593 | transport protein USO | 1.62e+01 |
| 38 | 46.9 | 265 | 2 G72340 | conserved hypothetical | 2.42e+01 |
| 39 | 46.9 | 302 | 2 T02480 | sec13-related protein | 2.42e+01 |
| 40 | 46.9 | 380 | 2 S49116 | hypothetical protein | 2.42e+01 |
| 41 | 46.9 | 444 | 1 WHRTW | tryptophan 5-monooxyg | 2.42e+01 |
| 42 | 46.9 | 646 | 2 B70396 | histidine kinase sens | 2.42e+01 |
| 43 | 46.9 | 836 | 2 A69550 | hypothetical protein | 2.42e+01 |
| 44 | 46.9 | 988 | 1 DJVZEP | DNA-directed DNA poly | 2.42e+01 |
| 45 | 46.9 | 1986 | 2 S28353 | probable polyketide s | 2.42e+01 |

ALIGNMENTS

RESULT 1
ENTRY IJHUG3 #type complete
TITLE desmoglein 3 precursor - human
ALTERNATE_NAMES pemphigus vulgaris antigen
ORGANISM #formal name Homo sapiens #common name man
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS
A41088
REFERENCE
#authors Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
#journal Cell (1991) 67:869-877
#title Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion.

#cross-references W01D:92069753
#accession A41088
#molecule_type mRNA
#residues 1-999 #label AMA
#cross-references GB:M76482; NID:g190751; PIDN:AAA60230.1; PID:g190754

GENETICS

#gene GDB:DSG3
#cross-references GDB:134030; OMIM:169615
#map_position 18q12.1-18q12.2
CLASSIFICATION
#superfamily cadherin: cadherin repeat homology
KEYWORDS
calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE

50-615
52-157
160-267
270-333
390-495
496-588
616-639
640-999
910-938
937-966
110,180,545
#domain signal sequence #status predicted #label SIG\
#domain propeptide #status predicted #label PRO\
#product desmoglein homolog #status predicted #label MAT\
#domain extracellular #status predicted #label EXT\
#domain cadherin repeat homology #label CR1\
#domain cadherin repeat homology #label CR2\
#domain cadherin repeat homology #label CR3\
#domain cadherin repeat homology #label CR4\
#domain cadherin repeat homology #label CR5\
#domain transmembrane #status predicted #label TMN\
#domain intracellular #status predicted #label INT\
#domain desmoglein repeat #label DG1\
#domain desmoglein repeat #label DG2\
#binding site carbohydrate (Asn) (covalent) #status predicted
#length 999 #molecular-weight 107502 #checksum 8311

SUMMARY
Query Match 100.0%; Score 113; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 1.81e-12;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 251 CECNIKVDVNDNP 265
 OY 1 CECNIKVDVNDNP 15

RESULT 2

ENTRY 1JBOG1 #type complete
 TITLE desmoglein 1 precursor - bovine
 ALTERNATE_NAMES desmoglein BDGM
 ORGANISM #formal_name Bos primigenius laurus #common_name cattle
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS S14603: A38872: A37785: S38721: A48173: S24412

REFERENCE
 #authors Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #submission Submitted to the EMBL Data Library, March 1991
 #description Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.

#accession S14603
 #molecule-type mRNA
 #residues 1-1043 #label KO2

REFERENCE
 #cross-references EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307

#authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.

#journal Eur. J. Cell Biol. (1991) 55:200-208

#title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.

#cross-references MUID:92037656

#accession A38872
 #molecule-type mRNA
 #residues 1-87;968-1043 #label KO2

REFERENCE
 #cross-references GB:S64268; GB:S64270

#authors Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowlin, P.

#journal Biochem. Biophys. Res. Commun. (1990) 173:1224-1230

#title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.

#cross-references MUID:91097553

#accession A37785
 #molecule-type mRNA
 #residues 44-123, 'V', 125-493 #label GOO

REFERENCE
 #cross-references GB:M58165; NID:g162966; PIDN:AAA62709.1; PID:g552318

#authors Zimbelmann, R.

#submission Submitted to the EMBL Data Library, February 1991

#accession S38721
 #molecule-type mRNA
 #residues 44-1043 #label ZIM

REFERENCE
 #cross-references EMBL:X57784; NID:g436061; PIDN:CAA0930.1; PID:g436062

#authors Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.

#journal Eur. J. Cell Biol. (1990) 53:1-12

#title Identification of desmoglein, a constitutive desmosomal glycoprotein, as a member of the cadherin family of cell adhesion molecules.

#cross-references MUID:91168965

#accession A48173
 #molecule-type mRNA
 #residues 44-1001, 'AOPPSAT' #label KO3

REFERENCE
 #cross-references GB:X57784

#note This sequence has been revised in references A38872 and S38721

GENETICS DSG1

CLASSIFICATION #superfamily cadherin; cadherin repeat homology

KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;

FEATURE

1-23
 24-49
 50-1043
 50-548
 52-157
 160-269
 272-385
 392-491
 549-574
 575-1043
 846-875
 876-905
 906-933
 934-962
 963-1012
 110

transmembrane protein

#domain signal sequence #status predicted #label SIG\

#domain propeptide #status predicted #label PRO\

#product desmoglein #status predicted #label MAT\

#domain extracellular #status predicted #label EXT\

#domain cadherin repeat homology #label CR1\

#domain cadherin repeat homology #label CR2\

#domain cadherin repeat homology #label CR3\

#domain cadherin repeat homology #label CR4\

#domain transmembrane #status predicted #label TM\

#domain intracellular #status predicted #label INT\

#domain desmoglein repeat #label DG1\

#domain desmoglein repeat #label DG2\

#domain desmoglein repeat #label DG3\

#domain desmoglein repeat #label DG4\

#region glycine/serine-rich\

#binding-site carbohydrate (Asn) (covalent) #status experimental\

#binding-site carbohydrate (Asn) (covalent) #status predicted

#length 1043 #molecular-weight 112242 #checksum 6897

SUMMARY

Query Match 78.8%; Score 89; DB 1; Length 1043;
 Best Local Similarity 80.0%; Pred. No. 9,39e-07;
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 253 CECNIKVDVNDNP 267
 OY 1 CECNIKVDVNDNP 15

RESULT 3

ENTRY 1JHUG1 #type complete
 TITLE desmoglein 1 precursor - human
 ALTERNATE_NAMES desmosomal glycoprotein I
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS S16906: A39706; A61254; A61279; S16158

REFERENCE
 #authors Buxton, R.S.

#submission Submitted to the EMBL Data Library, November 1990

#accession S16906
 #molecule-type mRNA
 #residues 1-1049 #label BUX

REFERENCE
 #cross-references EMBL:X56654; NID:g30505; PIDN:CAA39976.1; PID:g30506

#authors Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arnemann, J.; Rutman, A.J.; Pidelley, S.C.; Watt, F.M.; Rees, D.A.; Buxton, R.S.; Magee, A.I.

#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4796-4800

#title Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.

#cross-references MUID:91271279

#accession A39706
 #molecule-type mRNA
 #residues 24-1049 #label WHE

REFERENCE
 #cross-references GB:X56654

#authors Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.

#journal J. Cell Sci. (1991) 99:809-821

#title Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.

#cross-references MUID:92121251

#accession A61254
 #molecule-type mRNA
 #residues 26-1049 #label NIL

REFERENCE
 #cross-references A61279

#authors Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee, A.I.

```

#Journal      Biochem. Soc. Trans. (1991) 19:1060-1064
#Title        Desmosomal glycoproteins I, II and III: novel members of the
               cadherin superfamily.
#cross-references WUID:92175187
#accession    A61279
#status       not compared with conceptual translation
#molecule_type mRNA
#residues     1-55 ##label WH3

GENETICS
#gene         GDB:DSG1
#cross-references GDB:126563; OMIM:125670
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin repeat homology
                calcium binding; cell adhesion; duplication; glycoprotein;
                transmembrane protein
FEATURES
1-23          #domain signal sequence #status predicted #label SIG\
24-49         #domain propeptide #status predicted #label PRO\
50-1049       #product desmoglein #status predicted #label M\
50-548        #domain extracellular #status predicted #label EXT\
52-157        #domain cadherin repeat homology #label CR1\
160-269       #domain cadherin repeat homology #label CR2\
272-385       #domain cadherin repeat homology #label CR3\
393-493       #domain cadherin repeat homology #label CR4\
509-530       #region serine/threonine-rich\
549-569       #domain transmembrane #status predicted #label TM\
572-1049      #domain intracellular #status predicted #label INT\
840-869       #domain desmoglein repeat #label DG1\
870-899       #domain desmoglein repeat #label DG2\
900-927       #domain desmoglein repeat #label DG3\
928-956       #domain desmoglein repeat #label DG4\
969-1019      #region glycine/serine-rich\
110,180       #binding-site carboxylate (Asn) (covalent) #status
               predicted

SUMMARY
#length 1049 #molecular-weight 113715 #checksum 4482

Query Match      78.8% Score 89: DB 1: Length 1049:
Best Local Similarity 80.0% Pred. No. 9,396-07:
Matches 12: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

Db 253 CECN1KVDVNDNF 267
Oy 1 CECN1KVDVNDNF 15

RESULT 4
ENTRY      S62588 #type complete
TITLE      hypothetical protein SPAC21E11.03c - fission yeast
            (Schizosaccharomyces pombe)
ORGANISM   #formal_name Schizosaccharomyces pombe
            16-May-1996 #sequence_revision 13-Mar-1997 #text_change
            31-Oct-1997
DATE       S62588
ACCESSIONS S62588
REFERENCE  S62586
            McLean, J.; Harris, D.
            #authors
            #submission submitted to the EMBL Data Library, November 1995
            #accession S62588
            #status Preliminary
            #molecule_type DNA
            #residues 1-171 ##label MCL
            #cross-references EMBL:Z67999; NID:g1067216; PID:g1067219

GENETICS
#map_position 1L
CLASSIFICATION #superfamily fos/jun DNA-binding domain homology
FEATURES
5-45          #domain fos/jun DNA-binding domain homology #label FJD
            #length 171 #molecular-weight 19348 #checksum 9337
SUMMARY
Query Match      58.4% Score 66: DB 2: Length 171:
Best Local Similarity 35.7% Pred. No. 8,666-02:
Matches 5: Conservative 6: Mismatches 3: Indels 0: Gaps 0:

Db 76 COCSVKIRSVLTDF 89

```

```

Oy 1 CECN1KVDVNDNF 14

RESULT 5
ENTRY      H64448 #type complete
TITLE      polyferredoxin - Methanococcus jannaschii
ORGANISM   #formal_name Methanococcus jannaschii
            13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
            11-Jun-1999
DATE       H64448
ACCESSIONS H64448
REFERENCE  H64300
            Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
            R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,
            R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
            Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
            Kirsness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
            Scott, J.L.; Geoghegan, N.S.M.; Meldrum, J.F.; Fuhrmann,
            J.L.; Nguyen, D.; Uitterlinden, T.R.; Kelley, J.M.; Peterson,
            J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
            K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
            H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
            Science (1996) 273:1058-1073
            #title Complete genome sequence of the methanogenic archaeon,
            Methanococcus jannaschii.
            #cross-references WUID:9637999
            #accession H64448
            #status Preliminary: nucleic acid sequence not shown:
            translation not shown

#molecule_type DNA
#residues 1-394 ##label BUL
#cross-references GB:U67360; GB:U77117; NID:g1591813; PIDN:AA99195.1;
            #cross-references PID:g1591821; TIGR:M01193; PID:g1511192

GENETICS
#map_position FOR134591-113575
CLASSIFICATION #superfamily polyferredoxin 6x2(4Fe-4S); ferredoxin 2(4Fe-4S)
                homology
FEATURES
6-34        #domain ferredoxin 2(4Fe-4S) homology #label FER1\
66-119      #domain ferredoxin 2(4Fe-4S) homology #label FER2\
135-189     #domain ferredoxin 2(4Fe-4S) homology #label FER3\
203-258     #domain ferredoxin 2(4Fe-4S) homology #label FER4\
272-324     #domain ferredoxin 2(4Fe-4S) homology #label FER5\
336-389     #domain ferredoxin 2(4Fe-4S) homology #label FER6\
SUMMARY
#length 394 #molecular-weight 43342 #checksum 4356

Query Match      55.8% Score 63: DB 2: Length 394:
Best Local Similarity 53.8% Pred. No. 3,406-01:
Matches 7: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

Db 83 CP1G1KVDVNDNF 95
Oy 3 CN1KVDVNDNF 15

RESULT 6
ENTRY      I1B0DD #type fragment
TITLE      desmocollin 2b precursor - bovine (fragment)
            epithelial type 2 desmocollin subform II
ORGANISM   #formal_name Bos primigenius taurus #common_name cattle
            30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
            22-Jun-1999
DATE       B41799
ACCESSIONS B41799
REFERENCE  A41799
            Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Troyanovsky,
            R.; Franke, W.W.
            #authors
            #journal Proc. Natl. Acad. Sci. U.S.A. (1992) 89:353-357
            #title Complexity and expression patterns of the desmosomal
            cadherins.
            #cross-references WUID:92108053
            #accession B41799
            #molecule_type mRNA
            #residues 1-809 ##label KOC

```

```

##cross-references GB:M81190. NID:9163757; PIDN:AAA30783.1; PID:9163755
##experimental_source muzzle
##note
##note
264-Gln and 333-Gln were also found

GENETICS
#gene
#accession
#molecule_type mRNA
#residues
##cross-references GB:M81190. NID:9163757; PIDN:AAA30783.1; PID:9163755
##experimental_source muzzle
##note
##note
264-Gln and 333-Gln were also found

CLASSIFICATION
#superfamily cadherin: cadherin repeat homology
#alternative splicing: calcium binding; cell adhesion;
#duplication: glycoprotein; transmembrane protein

FEATURE
1-89
90-809 #domain propeptide #status predicted #label PRO\
90-645 #product desmocollin 2b #status predicted #label MAT\
92-197 #domain extracellular #status predicted #label EXT\
200-309 #domain cadherin repeat homology #label CR1\
312-423 #domain cadherin repeat homology #label CR2\
426-526 #domain cadherin repeat homology #label CR3\
527-634 #domain cadherin repeat homology #label CR4\
646-672 #domain cadherin repeat homology #label CR5\
673-863 #domain transmembrane #status predicted #label TMN\
120,346,495,579 #domain intracellular #status predicted #label INT\
#binding_site carbohydrate (Asn) (covalent) #status
predicted

SUMMARY
#length 809 #checksum 9747

Query Match 53.1%; Score 60; DB 1; Length 809;
Best Local Similarity 61.5%; Pred. No. 1.28e+00;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 295 CINIENVNDMLP 307
1 1 1 1 1 1 1 1 1 1
OY 3 CNIKVKVDNDFP 15

RESULT 7
ENTRY IJBOBC #type fragment
TITLE desmocollin 2a precursor - bovine (fragment)
ALTERNATE_NAMES epithelial type 2 desmocollin subform I
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
22-Jun-1999

ACCESSIONS
REFERENCE A41799
#authors Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Troyanovsky,
R.; Franke, W.W.
#journal Proc. Natl. Acad. Sci. U.S.A. (1992) 89:353-357
#title Complexity and expression patterns of the desmosomal
cadherins.
#cross-references MUID:92108053
#accession A41799

GENETICS
#molecule_type mRNA
#residues 1-863 #label KOC
##cross-references GB:M81190. NID:9163757; PIDN:AAA30783.1; PID:9163755
##experimental_source muzzle
##note
##note
264-Gln and 333-Gln were also found

CLASSIFICATION
#superfamily cadherin: cadherin repeat homology
#alternative splicing: calcium binding; cell adhesion;
#duplication: glycoprotein; transmembrane
protein

FEATURE
1-89
90-645 #domain propeptide #status predicted #label PRO\
90-645 #product desmocollin 2a #status predicted #label MAT\
92-197 #domain extracellular #status predicted #label EXT\
200-309 #domain cadherin repeat homology #label CR1\
312-423 #domain cadherin repeat homology #label CR2\
426-526 #domain cadherin repeat homology #label CR3\
527-634 #domain cadherin repeat homology #label CR4\
646-672 #domain cadherin repeat homology #label CR5\
673-863 #domain transmembrane #status predicted #label TMN\
120,346,495,579 #domain intracellular #status predicted #label INT\
#binding_site carbohydrate (Asn) (covalent) #status
predicted

```

```

826 #binding_site phosphate (Ser) (covalent) #status
predicted
SUMMARY
#length 863 #checksum 1299

Query Match 53.1%; Score 60; DB 1; Length 863;
Best Local Similarity 61.5%; Pred. No. 1.28e+00;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 295 CINIENVNDMLP 307
1 1 1 1 1 1 1 1 1 1
OY 3 CNIKVKVDNDFP 15

RESULT 8
ENTRY I45858 #type complete
TITLE desmocollin - bovine
ALTERNATE_NAMES #formal_name Bos primigenius taurus #common_name cattle
DATE 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change
20-Aug-1999

ACCESSIONS
REFERENCE I45858
#authors Yue, K.K.; Holton, J.L.; Clarke, J.P.; Hyam, J.L.; Hashimoto,
T.; Chidgey, M.A.; Garrod, D.R.
#journal J. Cell Sci. (1995) 108:2163-2173
#title Characterisation of a desmocollin isoform (bovine DSC3)
exclusively expressed in lower layers of stratified
epithelia.
#cross-references MUID:95403557
#accession I45858
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-896 #label YUE
##cross-references GB:L33774; NID:9914820; PIDN:MAC41625.1; PID:9914821

GENETICS
#gene
#accession
#molecule_type mRNA
#residues 1-896 #label YUE
##cross-references GB:L33774; NID:9914820; PIDN:MAC41625.1; PID:9914821

CLASSIFICATION
#superfamily cadherin: cadherin repeat homology
#duplication: glycoprotein; transmembrane
protein

SUMMARY
#length 896 #molecular_weight 98687 #checksum 6602

Query Match 53.1%; Score 60; DB 2; Length 896;
Best Local Similarity 69.2%; Pred. No. 1.28e+00;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 340 CIIYKSDNMLP 352
1 1 1 1 1 1 1 1 1 1
OY 3 CNIKVKVDNDFP 15

RESULT 9
ENTRY T16176 #type complete
TITLE hypothetical protein F26F4.10 - Caenorhabditis elegans
ALTERNATE_NAMES #formal_name Caenorhabditis elegans
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
20-Sep-1999

ACCESSIONS
REFERENCE T16176
#authors Fulton, L.
#journal Submitted to the EMBL Data Library, March 1996
#title The sequence of C. elegans cosmid F26F4.
#accession T16176
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-709 #label FUL
##cross-references EMBL:U12964; NID:91213452; PID:9529208;
PIDN:AAA91224.1; CESP:F26F4.10

GENETICS
#experimental_source strain Bristol N2

FEATURE
#gene
#accession
#molecule_type DNA
#residues 1-709 #label FUL
##cross-references EMBL:U12964; NID:91213452; PID:9529208;
PIDN:AAA91224.1; CESP:F26F4.10

SUMMARY
Query Match 52.2%; Score 59; DB 2; Length 709;

```

Best Local Similarity 54.5%; Pred. No. 1.98e+00;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 375 DIKIKVGESEF 385
OY 4 NIKVKDVNDNF 14

RESULT 10
ENTRY T15276 #type complete
TITLE hypothetical protein R10F2.1 - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

ACCESSIONS T15276
REFERENCE 218320
#authors Du, Z.; Gattung, S.
#submision submitted to the EMBL Data Library, May 1997
#description The sequence of C. elegans cosmid R10F2.
#accession T15276
#status preliminary; translated from GB/EMBL/DBJ

GENETICS
#molecule-type DNA
#residues 1-2163 #label DUZ
#cross-references EMBL:AF003388; NID:q2088850; PID:q2088852;
#experimental_source strain Bristol N2; clone R10F2
PIDB:GN00021; CESP:R10F2.1

GENE CESP:R10F2.1
#map_position 3
#introns 34/1; 114/3; 735/3; 1552/2; 1800/3; 1865/3; 1911/3; 2026/3;
2108/2; 2150/3

SUMMARY #length 2163 #molecular-weight 238609 #checksum 7114

Query Match 52.2%; Score 59; DB 2; Length 2163;
Best Local Similarity 53.3%; Pred. No. 1.98e+00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 1629 CKCHVIVLNDNDLP 1643
OY 1 CECNIKVKDVNDNF 15

RESULT 11
ENTRY S62484 #type complete
TITLE hypothetical protein SPAC4G8.07c - fission yeast
ORGANISM (Schizosaccharomyces pombe)
#formal_name Schizosaccharomyces pombe
DATE 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 31-Oct-1997

ACCESSIONS S62484
REFERENCE S62445
#authors Badcock, K.; Churcher, C.M.
#submision submitted to the EMBL Data Library, October 1995
#accession S62484
#status preliminary

GENETICS
#molecule-type DNA
#residues 1-527 #label BAD
#cross-references EMBL:z56276; NID:g1022345; PID:g1022352

SUMMARY #map_position 1L
#introns 125/1; 158/2; 437/3; 520/3
#length 527 #molecular-weight 59613 #checksum 8195

Query Match 51.3%; Score 58; DB 2; Length 527;
Best Local Similarity 42.9%; Pred. No. 3.05e+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 275 ECIPIATKTINEEYP 288
OY 2 ECNIKVKDVNDNF 15

RESULT 12

ENTRY S55396 #type complete
TITLE LI-cadherin - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 10-Sep-1997

ACCESSIONS S55396
REFERENCE S55396
#authors Boettlinger, A.; Kreft, B.; Flegler, C.; Dlouhy, B.; Berndorff, D.; Goessens, R.; Tauber, R.
#submision submitted to the EMBL Data Library, December 1994
#description Molecular cloning of human LI-cadherin: evidence for a novel type of cadherin within the cadherin superfamily.

ACCESSION S55396
#status preliminary
#molecule-type mRNA
#residues 1-832 #label BOE
#cross-references EMBL:X83228; NID:g854174; PID:g854175
CLASSIFICATION #superfamily cadherin repeat homology
FEATURE 455-566

SUMMARY #domain cadherin repeat homology #label CR3
#length 832 #molecular-weight 92207 #checksum 9645

Query Match 51.3%; Score 58; DB 2; Length 832;
Best Local Similarity 64.3%; Pred. No. 3.05e+00;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 325 EHKVKVDNDNF 338
OY 2 ECNIKVKDVNDNF 15

RESULT 13
ENTRY S51622 #type complete
TITLE cut3 protein - fission yeast (Schizosaccharomyces pombe)
ORGANISM #formal_name Schizosaccharomyces pombe
DATE 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 26-Aug-1999

ACCESSIONS S51622
REFERENCE S51622
#authors Saka, Y.; Sutani, T.; Yamashita, Y.; Satoh, S.; Takeuchi, M.; Nakaseko, Y.; Yanagida, M.
#journal EMBO J. (1994) 13:4938-4952
#title Fission yeast cut3 and cut14, members of a ubiquitous protein family, are required for chromosome condensation and segregation in mitosis.
#cross-references MUID:95045386

CLASSIFICATION
#molecule-type DNA
#residues 1-1324 #label SAK
#cross-references EMBL:D30788; NID:g577659; PID:d1007025; PID:g603501
SUMMARY #superfamily conserved hypothetical P115 protein
#length 1324 #molecular-weight 150593 #checksum 3330

Query Match 51.3%; Score 58; DB 2; Length 1324;
Best Local Similarity 42.9%; Pred. No. 3.05e+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 200 CDVEITFEVNSDF 213
OY 1 CECNIKVKDVNDNF 14

RESULT 14
ENTRY E70378 #type complete
TITLE DNA replication protein Dnac - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998

ACCESSIONS E70378
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.;

Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
#cross-references MIMD:98196666
#accession E70378
#status preliminary: nucleic acid sequence not shown;
translation not shown

##molecule-type DNA
##residues 1-235 #label AQF
##cross-references GB:AE00013; NID:g2983424; PID:g2983431; GB:AE000657
##experimental_source strain VFS

GENETICS

SUMMARY #gene dnac #length 235 #molecular-weight 26934 #checksum 1332

Query Match 50.4%; Score 57; DB 2; Length 235;
Best Local Similarity 63.6%; Pred. No. 4.66e+00;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 26 CECRFRKRDVN 36

OY 1 CECNIXVKDVN 11

RESULT 15

ENTRY T10664 #type complete
TITLE serine/threonine-specific protein kinase-like - Arabidopsis
thaliana

ALTERNATE_NAMES protein F6E21.20
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress

DATE 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change
20-Sep-1999

ACCESSIONS T10664
REFERENCE 216533

#authors

Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream,
M.A.; Barrell, B.G.; Bancroft, I.; Mewes, H.W.; Mayer,
K.F.X.; Lemcke, K.; Schellier, C.

#submission submitted to the Protein Sequence Database, June 1999

#accession T10664

#status preliminary

##molecule-type DNA

##residues 1-656 #label BEV

##cross-references EMBL:AL049914; GSPDB:GN00062; ATSP:F6E21.20

##experimental_source cultivar Columbia; BAC clone F6E21

GENETICS

#gene ATSP:F6E21.20

#map_position 4

#introns 206/L1: 243/L1

SUMMARY #length 656 #molecular-weight 73503 #checksum 8670

Query Match 50.4%; Score 57; DB 2; Length 656;
Best Local Similarity 50.0%; Pred. No. 4.66e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 198 CECNLGYKDI 209

OY 1 CECNIXVKDVND 12

Search completed: Sat May 13 07:21:44 2000
Job time: 8 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 07:17:05 2000; Maspar time 66.90 Seconds
6.829 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pep
Perfect Score: 113
Sequence: 1 CECNIKVKVDVNDNFP 15

Scoring table:
PAM 150
Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 26.676; Variance 31.455; scale 0.848

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|------------|------------------------|-----------|
| 1 | 113 | 100.0 | 999 | 1 | DSG3_HUMAN | DESMOGLN 3 PRECURSOR | 5.22e-15 |
| 2 | 89 | 78.8 | 1043 | 1 | DSG1_BOVIN | DESMOGLN 1 PRECURSOR | 2.23e-08 |
| 3 | 89 | 78.8 | 1043 | 1 | DSG1_HUMAN | DESMOGLN 1 PRECURSOR | 2.23e-08 |
| 4 | 66 | 58.4 | 171 | 1 | PCRI_SCHPO | TRANSFERRIN FACTOR P | 1.19e-02 |
| 5 | 62 | 54.9 | 501 | 1 | YONT_CAEEL | PUTATIVE G PROTEIN-COU | 9.47e-02 |
| 6 | 60 | 53.1 | 863 | 1 | DSC3_BOVIN | DESMOGLN 2A/2B PREC | 2.60e-01 |
| 7 | 60 | 53.1 | 896 | 1 | DSC3_BOVIN | DESMOGLN 3A/3B PREC | 2.60e-01 |
| 8 | 59 | 52.2 | 709 | 1 | SYR_CAEEL | PROBABLE ARGINYL-TRNA | 4.27e-01 |
| 9 | 59 | 52.2 | 780 | 1 | OBP_HSV6 | REPLICATION ORIGIN BIN | 4.27e-01 |
| 10 | 59 | 52.2 | 780 | 1 | OBP_HSV6 | REPLICATION ORIGIN BIN | 4.27e-01 |
| 11 | 59 | 52.2 | 902 | 1 | DSC2_MOUSE | DESMOGLN 2A/2B PREC | 6.96e-01 |
| 12 | 58 | 51.3 | 527 | 1 | VAD7_SCHPO | HYPOTHETICAL 59.6 KD P | 6.96e-01 |
| 13 | 58 | 51.3 | 1324 | 1 | CUT3_SCHPO | CHROMOSOME SEGREGATION | 1.13e+00 |
| 14 | 57 | 50.4 | 783 | 1 | CAD5_MOUSE | VASCULAR ENDOTHELIAL-C | 1.13e+00 |
| 15 | 57 | 50.4 | 784 | 1 | CAD5_MOUSE | VASCULAR ENDOTHELIAL-C | 1.13e+00 |
| 16 | 57 | 50.4 | 901 | 1 | DSC2_HUMAN | DESMOGLN 2A/2B PREC | 1.82e+00 |
| 17 | 56 | 49.6 | 413 | 1 | DHNO_AGR7 | D-NOPALINE DEHYDROGENA | 1.82e+00 |
| 18 | 56 | 49.6 | 782 | 1 | CAD5_PIG | VASCULAR ENDOTHELIAL-C | 1.82e+00 |
| 19 | 56 | 49.6 | 1807 | 1 | ITB4_RAT | INTEGRIN BETA-4 PRECUR | 1.82e+00 |
| 20 | 56 | 49.6 | 1875 | 1 | ITB4_HUMAN | INTEGRIN BETA-4 PRECUR | 1.82e+00 |
| 21 | 54 | 47.8 | 249 | 1 | Y927_METUA | HYPOTHETICAL PROTEIN M | 4.66e+00 |
| 22 | 54 | 47.8 | 670 | 1 | TBUD_BURPI | PHENOL 2-MONOXYGENASE | 4.66e+00 |
| 23 | 54 | 47.8 | 1790 | 1 | USOL_YEAST | INTRACELLULAR PROTEIN | 4.66e+00 |

ALIGNMENTS

| RESULT ID | 1 | STANDARD: | PRT: | 999 AA. |
|-----------|--|-----------|------|---------|
| AC | DSG3_HUMAN | | | |
| AC | P32926 | | | |
| DT | 01-OCT-1993 (Rel. 27, Created) | | | |
| DT | 01-OCT-1993 (Rel. 27, Last sequence update) | | | |
| DT | 01-NOV-1997 (Rel. 35, Last annotation update) | | | |
| DE | DESMOGLN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA). | | | |
| GN | DSG3. | | | |
| OS | Homo sapiens (Human). | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; | | | |
| CC | Eutheria; Primates; Catarrhini; Hominoidea; Homo. | | | |
| CC | [1] | | | |
| CC | SEQUENCE FROM N.A. | | | |
| CC | Amagaki M., Klaus-Kovtun V., Stanley J.R.; | | | |
| CC | "Autoantibodies against a novel epithelial cadherin in pemphigus | | | |
| CC | vulgaris, a disease of cell adhesion."; | | | |
| CC | Cell 67:869-877(1991). | | | |
| CC | - FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. | | | |
| CC | - INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE | | | |
| CC | FILAMENTS MEDIATING CELL-CELL ADHESION. | | | |
| CC | - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. | | | |
| CC | - TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND | | | |
| CC | CARCINOMAS. | | | |
| CC | - DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS | | | |
| CC | (POTENTIAL). | | | |
| CC | - DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN | | | |
| CC | CONDITION IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE | | | |
| CC | LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES | | | |
| CC | AGAINST DSG3. | | | |
| CC | - SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation. | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL: M76482; AAA60230.1; - | | | |
| DR | PIR: A41088; ITHUG3. | | | |
| DR | HSSP: P09803; 1EDH. | | | |
| DR | MM: 169615; - | | | |
| DR | PROSITE: PS00232; CADHERIN. 3. | | | |
| DR | PFAM: PF00028; cadherin. 4. | | | |
| CC | Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein; | | | |

Calcium-binding: Repeat.

CC SIGNAL 1 23 POTENTIAL.

CC PROPEP 24 49 POTENTIAL.

CC CHAIN 50 999 DESMOGLEIN 3.

CC DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).

CC TRANSMEM 616 640 POTENTIAL.

CC DOMAIN 641 999 CYTOPLASMIC (POTENTIAL).

CC REPEAT 50 158 CADHERIN 1.

CC REPEAT 159 268 CADHERIN 2.

CC REPEAT 269 383 CADHERIN 3.

CC REPEAT 386 499 CADHERIN 4.

CC REPEAT 910 935 DESMOGLEIN REPEAT 1.

CC REPEAT 936 966 DESMOGLEIN REPEAT 2.

CC CARBOHYD 110 110 POTENTIAL.

CC CARBOHYD 180 180 POTENTIAL.

CC CARBOHYD 459 459 POTENTIAL.

CC CARBOHYD 545 545 POTENTIAL.

CC SEQUENCE 999 AA: 107503 MW: 60479DD46AC219A1 CRC64;

Query Match 100.0%; Score 113; DB 1; Length 999;

Best Local Similarity 100.0%; Pred. No. 5,22e-15;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2b 251 CECNIRKVDVNDNFP 265

2y 1 CECNIRKVDVNDNFP 15

RESULT 2 STANDARD; PRT: 1043 AA.

AC 003763;

CC 01-OCT-1993 (Rel. 27, Created)

CC 01-OCT-1993 (Rel. 27, Last sequence update)

CC 01-NOV-1997 (Rel. 35, Last annotation update)

CC DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DGI).

CC DSG1.

CC Bos taurus (Bovine).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

CC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;

CC Bovinae; Bos.

CC (1)

CC SEQUENCE FROM N.A.

CC TISSUE=MUZZLE EPITHELIUM;

CC Koch P.J., Goldschmidt M.D., Zimbelmann R., Franke W.W.;

CC Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.

CC [2]

CC SEQUENCE OF 44-1043 FROM N.A.

CC TISSUE=MUZZLE EPITHELIUM;

CC MEDLINE: 91168965.

CC Koch P.J., Walsh M.J., Schmelz M., Goldschmidt M.D.,

CC Zimbelmann R., Franke W.W.;

CC "Identification of desmoglein, a constitutive desmosomal

CC glycoprotein, as a member of the cadherin family of cell adhesion

CC molecules.";

CC Eur. J. Cell Biol. 53:1-12(1990).

CC [3]

CC REVISIONS, AND SEQUENCE OF 101-123.

CC MEDLINE: 92037656.

CC Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R.,

CC Franke W.W.;

CC "Complete amino acid sequence of the epidermal desmoglein precursor

CC polypeptide and identification of a second type of desmoglein gene.";

CC Eur. J. Cell Biol. 55:200-208(1991).

CC [4]

CC SEQUENCE OF 44-493 FROM N.A.

CC MEDLINE: 91097553.

CC Goodwin L., Hill J.E., Raynor K., Raszi L., Manabe M., Cowin P.;

CC "Desmoglein shows extensive homology to the cadherin family of cell

CC adhesion molecules.";

CC Biochem. Biophys. Res. Commun. 173:1224-1230(1990).

CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.

CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE

CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.

CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X58466; CAA41380.1; -

CC EMBL: X57784; CAA40930.1; -

CC EMBL: M58165; AAA62709.1; -

CC PIR: S14603; IJBOG1.

CC HSSP: P09803; 1EDH.

CC PROSITE: P800232; CADHERIN; 2.

CC PFAM: PF00028; cadherin; 3.

CC Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;

CC Calcium-binding; Repeat.

CC SIGNAL 1 23 POTENTIAL.

CC PROPEP 24 49 POTENTIAL.

CC CHAIN 50 1043 EXTRACELLULAR (POTENTIAL).

CC TRANSMEM 549 573 POTENTIAL.

CC DOMAIN 574 1043 CYTOPLASMIC (POTENTIAL).

CC REPEAT 50 158 CADHERIN 1.

CC REPEAT 159 270 CADHERIN 2.

CC REPEAT 271 385 CADHERIN 3.

CC REPEAT 386 498 CADHERIN 4.

CC REPEAT 819 845 DESMOGLEIN REPEAT 1.

CC REPEAT 846 875 DESMOGLEIN REPEAT 2.

CC REPEAT 876 905 DESMOGLEIN REPEAT 3.

CC REPEAT 906 933 DESMOGLEIN REPEAT 4.

CC REPEAT 934 962 DESMOGLEIN REPEAT 5.

CC DOMAIN 963 1012 GLY/SER-RICH.

CC CARBOHYD 110 110 POTENTIAL.

CC CARBOHYD 180 180 POTENTIAL.

CC CARBOHYD 496 496 POTENTIAL.

CC CONFLICT 124 124 I -> V (IN REF. 4).

CC SEQUENCE 1043 AA: 112243 MW: ADE46133F8B77C11 CRC64;

Query Match 78.8%; Score 89; DB 1; Length 1043;

Best Local Similarity 80.0%; Pred. No. 2.23e-08;

Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 253 CECNIRKVDVNDNFP 267

Qy 1 CECNIRKVDVNDNFP 15

RESULT 3 STANDARD; PRT: 1049 AA.

AC 002413;

CC 01-OCT-1993 (Rel. 27, Created)

CC 01-OCT-1993 (Rel. 27, Last sequence update)

CC 01-NOV-1997 (Rel. 35, Last annotation update)

CC DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DGI).

CC DSG1.

CC Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

CC Eutheria; Primates; Catarrhini; Hominiidae; Homo.

CC (1)

CC SEQUENCE FROM N.A.

CC TISSUE=KERATINOCYTES;

CC MEDLINE: 91271279.

CC Wheeler G.N., Parker A.E., Thomas C.L., Ataliotis P., Poynter D.,

CC Arneimann J., Rutan A.V., Pidsley S.C., Watt F.M., Rees D.A.,

CC Buxton R.S., Magee A.I.;

CC "Desmosomal glycoprotein DGI, a component of intercellular desmosome

RT junctions, is related to the cadherin family of cell adhesion molecules.":

RT Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).

CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.

CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONGUE-LIKE ESOPHAGUS.

CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X56554; CAA39976.1; -

DR PIR: S16906; IJHUG1.

DR HSSP: P09803; IEDH.

DR MIM: 125670; -

DR PROSITE: PS00232; CADHERIN; 2.

DR PFAM: PF00028; cadherin; 4.

KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein; Calcium-binding; Repeat.

KM SIGNAL 1 23

FT PROPEP 24 49 POTENTIAL.

FT CHAIN 50 1049 DESMOGLEIN 1.

FT DOMAIN 50 545 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 546 570 POTENTIAL.

FT DOMAIN 571 1049 CYTOPLASMIC (POTENTIAL).

FT REPEAT 50 158 CADHERIN 1.

FT REPEAT 159 270 CADHERIN 2.

FT REPEAT 271 385 CADHERIN 3.

FT REPEAT 386 497 CADHERIN 4.

FT REPEAT 813 839 DESMOGLEIN REPEAT 1.

FT REPEAT 840 869 DESMOGLEIN REPEAT 2.

FT REPEAT 870 899 DESMOGLEIN REPEAT 3.

FT REPEAT 900 927 DESMOGLEIN REPEAT 4.

FT REPEAT 928 956 DESMOGLEIN REPEAT 5.

FT DOMAIN 969 1019 GLY/SER-RICH.

FT CARBOHYD 36 36 POTENTIAL.

FT CARBOHYD 110 110 POTENTIAL.

FT CARBOHYD 180 180 POTENTIAL.

SO SEQUENCE 1049 AA; 113715 MW; EE125655B9D6619 CRC64;

Query Match Score 89; DB 1; Length 1049;

Best Local Similarity 80.0%; Pred. No. 2,23e-08;

Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 253 CECNIIKIDVNDNIP 267

OY 1 CECNIIKIVKDVNDNF 15

RESULT 4

ID PCRI-SCHPO STANDARD: PRT: 171 AA.

AC 009926;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE TRANSCRIPTION FACTOR PCRI (TRANSCRIPTION FACTOR MTS2).

GN PCRI OR MTS2 OR SPAC21E1.03C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;

OC Schizosaccharomycetaceae; Schizosaccharomycetes.

CC [1]

RN SEQUENCE FROM N.A.

RP MEDLINE: 9610438.

RA Watanabe Y., Yamamoto M.;

RT "Schizosaccharomyces pombe pcrl+ encodes a CREB/ATF protein involved in regulation of gene expression for sexual development.":

RT Mol. Cell. Biol. 16:704-711(1996).

CC [2]

RP SEQUENCE FROM N.A.

RA Kon N., Krawchuk M.D., Warren B.G., Smith G.R., Wahls W.P.;

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-972;

RL McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;

RN Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.

RP CHARACTERIZATION.

RX MEDLINE: 95047325.

RA Wahls W.P., Smith G.R.;

RT "A heteromeric protein that binds to a meiotic homologous recombination hot spot: correlation of binding and hot spot activity."

RT Genes Dev. 8:1693-1702(1994).

CC -1- FUNCTION: INVOLVED IN REGULATION OF GENE EXPRESSION FOR SEXUAL DEVELOPMENT. BINDS AND ACTIVATES MEIOTIC RECOMBINATION HOT SPOT ADE6-M26.

CC -1- SUBUNIT: HETERODIMER OF PCRI/MTS2 AND ATF1/MTS1.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D63667; BAA0918.1; -

DR EMBL: U87870; ABA46991.1; -

DR EMBL: Z67999; CAA91968.1; -

DR HSSP: P05412; IFOS.

DR TRANSFAC: T01687; -

DR PROSITE: PS00036; BZIP-BASIC; 1.

DR PFAM: PF00170; bZIP; 1.

KW Transcription regulation; Activator; DNA-binding; Nuclear protein; Melosis.

FT DNA-BIND 12 32 BASIC MOTIF.

FT DOMAIN 42 66 LEUCINE-ZIPPER.

SO SEQUENCE 171 AA; 19348 MW; 9922FDDDEF150BDE CRC64;

Query Match Score 66; DB 1; Length 171;

Best Local Similarity 35.7%; Pred. No. 1.19e-02;

Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

DB 76 COCSVKIRSVLTDF 89

OY 1 CECNIIKIVKDVNDNF 14

RESULT 5

ID YONJ-CAEEL STANDARD: PRT: 501 AA.

AC 002213;

DT 15-FEB-2000 (Rel. 39, Created)

DT 15-FEB-2000 (Rel. 39, Last sequence update)

DT 15-FEB-2000 (Rel. 39, Last annotation update)

DE PUTATIVE G PROTEIN-COUPLED RECEPTOR C02D4.2.

GN C02D4.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;

OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

CC [1]

RN SEQUENCE FROM N.A.

RP STRAIN-BRISTOL N2;

RA Baynes C.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

```

CC -1- FUNCTION: PUTATIVE G-PROTEIN RECEPTOR.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC MOST SIMILAR TO INSECT OCTOPAMINE RECEPTORS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 281031; CAB02718.1; -.
DR MORPEP: C0204.2; CE07837.
DR PROSITE: PS00237; G-PROTEIN_RECEPTOR: 1.
DR PIR: P00001; 7tm_1: 2.
KM G-protein coupled receptor: Transmembrane; Glycoprotein.
FT DOMAIN 1 49 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 50 72 1 (POTENTIAL).
FT DOMAIN 73 82 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 83 104 2 (POTENTIAL).
FT DOMAIN 120 136 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 137 157 3 (POTENTIAL).
FT DOMAIN 158 177 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 178 200 4 (POTENTIAL).
FT DOMAIN 201 224 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 225 246 5 (POTENTIAL).
FT DOMAIN 247 373 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 374 395 6 (POTENTIAL).
FT DOMAIN 396 426 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 427 447 7 (POTENTIAL).
FT DOMAIN 448 487 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 493 496 POLY-ARG.
FT CARBOHYD 15 15 POTENTIAL.
FT DISULFID 134 212 BY SIMILARITY.
SQ SEQUENCE 501 AA: 56404 MW: 5184614065228F8 CRC64:

Query Match 54.9%: Score 62; DB 1; Length 501;
Best Local Similarity 54.5%: Pred. No. 9.47e-02;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 336 CAYKVKDVKED 346
   1:::|||||:
OY 3 CNIKVKVDNDN 13

RESULT 6
ID DSC2_BOVIN STANDARD: PRT: 863 AA.
AC P33545;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DESMOCOLLIN 2A/2B PRECURSOR (EPITHELIAL TYPE 2 DESMOCOLLIN)
DE (FRAGMENT).
GN DSC2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE EPITHELIUM;
RX MEDLINE: 92108053.
RA Koch P.J., Goldschmidt M.D., Zimbelmann R., Troyanovsky R.,
RA Franke W.W.;
RT "Complexity and expression patterns of the desmosomal cadherins.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:353-357(1992).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOTFORMS.

```

```

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: FORMS 2A (SHOWN HERE) AND 2B ARE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE.
CC -1- TISSUE SPECIFICITY: ESOPHAGUS AND ROMEN. WEAKLY IN EPITHELIA AND
CC CARDIAC MUSCLE.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M81190; AAA30783.1; -.
DR EMBL: M81190; AAA30783.1; -.
DR PIR: A41799; IUBODC.
DR PIR: B41799; IUBODC.
DR HSSP: P09803; ISUH.
DR PROSITE: PS00232; CADHERIN: 3.
DR PIR: P000028; cadherin: 5.
KM Cell adhesion: Glycoprotein; Transmembrane; Repeat;
KM Alternative splicing: Cytoskeleton; Calcium-binding.
FT NON_TER 1 1
FT PROP 1 89 POTENTIAL.
FT CHAIN 90 863 DESMOCOLLIN 2A/2B.
FT DOMAIN 90 644 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 645 665 POTENTIAL.
FT TRANSMEM 666 863 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 90 197 CADHERIN 1.
FT REPEAT 198 309 CADHERIN 1.
FT REPEAT 310 423 CADHERIN 2.
FT REPEAT 424 528 CADHERIN 3.
FT REPEAT 529 644 CADHERIN 4.
FT CARBOHYD 120 120 CADHERIN 5.
FT CARBOHYD 346 346 POTENTIAL.
FT CARBOHYD 495 495 POTENTIAL.
FT CARBOHYD 579 579 POTENTIAL.
FT VARIANT 264 264 K -> Q.
FT VARIANT 333 333 R -> D.
FT VARSPLIC 799 809 KVGQCDDNDTH -> ESIGHGLVKN (IN ISOFORM
FT VARSPLIC 810 863 MISSING (IN ISOFORM 2B).
SQ SEQUENCE 863 AA: 95874 MW: 6038540CA16727F2 CRC64:

Query Match 53.1%: Score 60; DB 1; Length 863;
Best Local Similarity 61.5%: Pred. No. 2.60e-01;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 295 CIINIEDVNDNP 307
   1:::|||||:
OY 3 CNIKVKVDNDNF 15

RESULT 7
ID DSC3_BOVIN STANDARD: PRT: 896 AA.
AC Q28060; Q28061; Q28176;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE DESMOCOLLIN 3A/3B PRECURSOR.
GN DSC3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95403557.
RA Yue K.K.M., Holton J.L., Clarke J.P., Hyam J.L.M., Hashimoto T.,

```

RA Chidgey M.A.J., Garrod D.R.:
 RT "Characterisation of a desmocollin isoform (bovine DSC3) exclusively
 RT expressed in lower layers of stratified epithelia.";
 RL J. Cell Sci. 108:2163-2173(1995).
 RN [2]
 RP SEQUENCE OF 686-814 FROM N.A.
 RC TISSUE-EPIDERMIS;
 RA MEDLINE: 94308280.
 RX Legan P.K., Yue K.K.M., Chidgey M.A.J., Holton J.L., Wilkinson R.W.,
 RA Garrod D.R.:
 RT "The bovine desmocollin family: a new gene and expression patterns
 RT reflecting epithelial cell proliferation and differentiation.";
 RL J. Cell Biol. 126:507-518(1994).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS. INVOLVED
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: FORMS 3A AND 3B ARE PRODUCED BY ALTERNATIVE
 CC SPLICING OF THE SAME GENE.
 CC -1- TISSUE SPECIFICITY: STRATIFIED EPITHELIA ONLY (EPIDERMIS, TONGUE,
 CC ESOPHAGUS AND RUMEN).
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L33774; AAC41625.1; -
 CC EMBL: L33774; AAC41626.1; -
 CC EMBL: X75783; CAA53427.1; -
 CC HSSP: P09803; ISUH.
 DR PRINTS: PR00205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN. 3.
 DR PFM: PF00028; cadherin. 5.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
 KW Alternative splicing; Cytoskeleton; Calcium-binding.
 FT SIGNAL 1 26
 FT PROPEP 27 134
 FT CHAIN 135 886
 FT DOMAIN 135 690
 FT TRANSMEM 691 711
 FT DOMAIN 712 896
 FT REPEAT 135 242
 FT REPEAT 243 354
 FT REPEAT 355 471
 FT REPEAT 472 579
 FT REPEAT 580 690
 FT CARBOHYD 165 165
 FT CARBOHYD 391 391
 FT CARBOHYD 546 546
 FT CARBOHYD 629 629
 FT VARSPLIC 832 839
 FT VARSPLIC 840 896
 FT CONFLICT 686 687
 SO SEQUENCE 896 AA; 99687 MW; 8CC0C30A63FB0BD4 CRC64;

Query Match 53.1%; Score 60; DB 1; Length 896;
 Best Local Similarity 69.2%; Pred. No. 2.50e-01;
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 340 CITTAKSDNDLP 352
 1 1 1 1 1 1 1 1 1 1
 QY 3 CNIKVQVNDNPF 15

RESULT 8
 ID SYRCAEEL STANDARD: PRT: 709 AA.
 AC Q19825;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROBABLE ARGINYL-TRNA SYNTHETASE (EC 6.1.1.19) (ARGININE--TRNA LIGASE)
 DE (ARGRS).
 GN F26f4.10.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Fullon L.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP + L-ARGININE + TRNA(ARC) -> AMP +
 CC PYROPHOSPHATE + L-ARGINYL-TRNA(ARC).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U12964; AAA91224.1; -
 CC MORPEP: F26f4.10; CE01258.
 DR PRINTS: PR01036; TRNASYNTHARG.
 DR PROSITE: PS00178; AA-TRNA-LIGASE_1; 1.
 DR PFM: PF00750; tRNA-synt.1d; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
 FT SIMILAR 252 263
 SO SEQUENCE 709 AA; 80477 MW; 175B7CAA6096B9D CRC64;

Query Match 52.2%; Score 59; DB 1; Length 709;
 Best Local Similarity 54.5%; Pred. No. 4.27e-01;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 375 DIKIKVGESEF 385
 1 1 1 1 1 1 1 1 1 1
 QY 4 NIKVQVNDNPF 14

RESULT 9
 ID OBP_HSVGU STANDARD: PRT: 780 AA.
 AC P52378;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE REPLICATION ORIGIN BINDING PROTEIN (ORBP).
 GN U73 OR 19R OR HDRPO.
 OS Herpes simplex virus (type 6 / strain Uganda-1102).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Roseolovirus.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 95266321.
 RA Campels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
 RA Martin M.E., Efsthaliou S., Citraxon M., Macculay H.A.;
 RT "The DNA sequence of human herpesvirus-6: structure, coding content,
 RT and genome evolution.";
 RL Virology 209:29-51(1995).
 RN (2)
 RP SEQUENCE OF 1-703 FROM N.A.
 RC MEDLINE: 95146942.
 RA Lawrence G.L., Nicholas J., Barrell B.G.;
 RT "Human herpesvirus 6 (strain U1102) encodes homologues of the
 RT conserved herpesvirus glycoprotein gp and the alphaherpesvirus

```

RT origin-binding protein."
RL J. Gen. Virol. 76:147-152(1995).
CC -1- FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION. BINDS THE ORIGIN
CC OF REPLICATION (ORI) (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL9,
CC EHV-1 53, AND VZV 51.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X83413; CA58365.1; -.
DR EMBL: M68963; AAA65581.1; -.
KM DNA replication; DNA-binding; ATP (POTENTIAL).
FT NP BIND 52
SQ SEQUENCE 780 AA: 89716 MW: 53F4B0E46C9BCB5F CRC64:

Query Match 52.2%; Score 59; DB 1; Length 780;
Best Local Similarity 53.3%; Pred. No. 4,27e-01;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

DB 424 CENSIKYDIVNGFP 438
OY 1 CENSIKYDIVNDNFP 15

RESULT 10
ID OBP_HSV62 STANDARD: PRT: 780 AA.
AC P52452.
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE REPLICATION ORIGIN BINDING PROTEIN (OBP).
OS U73 OR CHER.
GN Herpes simplex virus (type 6 / strain 229).
OC Viruses; dsDNA viruses; no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96195263.
RA Lindquist G.J., Inoue N., Allen R.D., Castelli J.W.,
RA Stamey F.R., Dambaugh T.R., O'Brian J.J., Danovich R.M.,
RA Fenkel N., Pellet P.E.;
RT "Restriction endonuclease mapping and molecular cloning of the human
RT herpesvirus 6 variant B strain 229 genome."
RL Arch. Virol. 141:367-379(1996).
CC -1- FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION. BINDS THE ORIGIN
CC OF REPLICATION (ORI) (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL9,
CC EHV-1 53, AND VZV 51.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L14772; AAB06356.1; -.
KM DNA replication; DNA-binding; ATP (POTENTIAL).
FT NP BIND 52
SQ SEQUENCE 780 AA: 1DE68504FA9624B6 CRC64:

Query Match 52.2%; Score 59; DB 1; Length 780;
Best Local Similarity 53.3%; Pred. No. 4,27e-01;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

DB 424 CENSIKYDIVNGFP 438

```

```

OY 1 CENSIKYDIVNDNFP 15

RESULT 11
ID DSC2_MOUSE STANDARD: PRT: 902 AA.
AC P55292; 064734;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DESMOCOLLIN 2A/2B PRECURSOR (EPITHELIAL TYPE 2 DESMOCOLLIN).
GN DSC2 OR DSC3
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-C57BL/6; TISSUE-EMBRYO;
RX MEDLINE: 95227276.
RA Lorimer J.E., Hall L.S., Clarke J.P., Collins J.E., Fleming T.P.,
RA Garrod D.R.;
RT "Cloning, sequence analysis and expression pattern of mouse
RT desmocollin 2 (DSC2), a cadherin-like adhesion molecule."
RL Mol. Membr. Biol. 11:229-236(1994).
RN [2]
RP SEQUENCE OF 344-637 FROM N.A.
RX STRAIN-C57BL/6; TISSUE-EMBRYO;
RX MEDLINE: 95048328.
RA Buxton R.S., Wheeler G.N., Pidsley S.C., Marsden M.D., Adams M.J.,
RA Jenkins N.A., Gilbert D.J., Copeland N.G.;
RT "Mouse desmocollin (Dsc3) and desmoglein (Dsg1) genes are closely
RT linked in the proximal region of chromosome 18."
RL Genomics 21:510-516(1994).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: FORMS 2A AND 2B ARE PRODUCED BY ALTERNATIVE
CC SPLICING OF THE SAME GENE.
CC -1- TISSUE SPECIFICITY: IN ALL EPITHELIA TESTED AND HEART.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L33779; AAA79177.1; -.
DR EMBL: L33779; AAA79176.1; -.
DR EMBL: X73885; CA52089.1; -.
DR HSP: P09803; IEDH.
DR MGD: MGI:103221; DSC2.
DR PRINTS: PR00205; CADHERIN.
DR PROSITE: PS00232; CADHERIN.
DR PFAM: PF00028; cadherin. 5.
KM Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KM Alternative splicing; Cytoskeleton; Calcium-binding.
FT SIGNAL 1
FT PROPEP 28 135
FT CHAIN 136 902
FT DOMAIN 136 694
FT TRANSMEM 695 715
FT DOMAIN 716 902
FT REPEAT 136 243
FT REPEAT 244 355
FT REPEAT 356 471
CADHERIN 3.

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X83930; CAAS8782.1; -
 DR HSSP: P09803; LEDH.
 DR MGD: MGI:105057; CDH5.
 DR PRINTS: PRO0205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN. 3.
 DR PIR: P01049; Cadherin. 5.
 DR PIR: P01049; Cadherin_C-term; 1.
 DR Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
 KW Calcium-binding; Repeat; Signal.
 FT SIGNAL 1 24
 FT PROPEP 25 45 POTENTIAL.
 FT CHAIN 46 783 VASCULAR ENDOTHELIAL-CADHERIN.
 FT DOMAIN 46 592 EXTRACELLULAR (POTENTIAL).
 FT TRANSEM 593 619 POTENTIAL.
 FT DOMAIN 620 783 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 46 148 CADHERIN 1.
 FT REPEAT 149 255 CADHERIN 2.
 FT REPEAT 256 370 CADHERIN 3.
 FT REPEAT 371 475 CADHERIN 4.
 FT REPEAT 476 592 CADHERIN 5.
 FT DOMAIN 737 752 SER-RICH.
 FT CARBOHYD 59 59 POTENTIAL.
 FT CARBOHYD 154 154 POTENTIAL.
 FT CARBOHYD 440 440 POTENTIAL.
 FT CARBOHYD 522 522 POTENTIAL.
 FT CARBOHYD 534 534 POTENTIAL.
 SO SEQUENCE 783 AA: 87847 MW: 240A82D663BCE71C CRC64:
 Query Match 50.4%; Score 57; DB 1; Length 783;
 Best Local Similarity 63.6%; Pred. No. 1.13e+00;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 243 IRLDINDNFP 253
 Oy 5 IKVRDNDNFP 15

RESULT 15
 ID CAD5_HUMAN STANDARD: PRT: 784 AA.
 AC P33151;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE VASCULAR ENDOTHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5)
 DE (7B4 ANTIGEN) (CD144 ANTIGEN).
 GN CDH5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ENDOTHELIAL CELLS:
 RX MEDLINE: 95353875.
 RA Breviario F., Cavada L., Corada M., Martin-Padura I., Navarro P.,
 RA Goley J., Introna M., Guino D., Lampugnani M.G., Dejana E.;
 RT "Functional properties of human vascular endothelial cadherin
 RT (7B4/cadherin-5), an endothelial-specific cadherin";
 RL Arterioscler. Thromb. Vasc. Biol. 15:1229-1239(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA;
 RX MEDLINE: 97362755.
 RA Ali J., Liao F., Martens E., Muller W.A.;
 RT "Vascular endothelial cadherin (VE-cadherin): cloning and role in
 RT endothelial cell-cell adhesion";
 RL Microcirculation 4:267-277(1997).

RN [3]
 RP SEQUENCE OF 5-784 FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 91283540.
 RA Suzuki S., Sano K., Tanihara H.;
 RT "Diversity of the cadherin family: evidence for eight new cadherins
 RT in nervous tissue.";
 RL Cell Regul. 2:261-270(1991).
 RN [4]
 RP PARTIAL SEQUENCE.
 RC TISSUE-ENDOTHELIAL CELLS;
 RX MEDLINE: 92394977.
 RA Lampugnani M.G., Resnati M., Rauterl M., Pigott R., Pisecane A.,
 RA Houen G., Roco L.P., Dejana E.;
 RT "A novel endothelial-specific membrane protein is a marker of
 RT cell-cell contacts.";
 RL J. Cell Biol. 118:1511-1522(1992).
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A
 CC IMPORTANT ROLE IN ENDOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE
 CC COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS. IT
 CC ASSOCIATES WITH ALPHA-CATENIN FORMING A LINK TO THE CYTOSKELETON.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. FOUND AT CELL-CELL
 CC BOUNDARIES AND PROBABLY AT CELL-MATRIX BOUNDARIES.
 CC -1- TISSUE SPECIFICITY: ENDOTHELIAL TISSUES AND BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
 CC -1- DATABASE: NAME-PRO: NOTE-CD guide CD144.htm.
 CC WWW: <http://www.ncbi.nlm.nih.gov/prov/cd/cd144.htm>.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X79981; CAAS6306.1; -
 DR EMBL: U84722; AAB41796.1; -
 DR EMBL: X59796; CAA42468.1; -
 DR PIR: S24305; ITHUC5.
 DR HSSP: P09803; LEDH.
 DR MIM: 601120.
 DR PRINTS: PRO0205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN. 3.
 DR PIR: P01049; Cadherin. 5.
 DR PIR: P01049; Cadherin_C-term; 1.
 KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
 KW Calcium-binding; Repeat; Signal.
 FT SIGNAL 1 25
 FT PROPEP 26 47 POTENTIAL.
 FT CHAIN 48 784 VASCULAR ENDOTHELIAL-CADHERIN.
 FT DOMAIN 48 593 EXTRACELLULAR (POTENTIAL).
 FT TRANSEM 594 620 POTENTIAL.
 FT DOMAIN 621 784 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 48 151 CADHERIN 1.
 FT REPEAT 152 258 CADHERIN 2.
 FT REPEAT 259 372 CADHERIN 3.
 FT REPEAT 373 477 CADHERIN 4.
 FT REPEAT 478 593 CADHERIN 5.
 FT DOMAIN 736 753 SER-RICH.
 FT CARBOHYD 61 61 POTENTIAL.
 FT CARBOHYD 112 112 POTENTIAL.
 FT CARBOHYD 157 157 POTENTIAL.
 FT CARBOHYD 362 362 POTENTIAL.
 FT CARBOHYD 442 442 POTENTIAL.
 FT CARBOHYD 523 523 POTENTIAL.
 FT CARBOHYD 535 535 POTENTIAL.
 FT CONFLICT 517 517 I -> T (IN REF. 1).
 SO SEQUENCE 784 AA: 87528 MW: 6DBEBFC4DA6899D1 CRC64:

Query Match 50.4% Score 57; DB 1; Length 784;
Best Local Similarity 81.8%; Pred. No. 1.13e+00;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 139 IKVHDVNDNMP 149
||| ||||| |
QY 5 IKVKDVNDNFP 15

Search completed: Sat May 13 07:18:23 2000
Job time : 78 secs.

M O D E R N
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_bp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:18:42 2000; Maspar time 139.87 seconds
7.436 Million cell updates/sec
Tabular output not generated.

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pep
Perfect Score: 113
Sequence: 1 CECNIKVKVDNDNFP 15

Scoring table:
PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database:

sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_orcanelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 25.749; Variance 32.815; scale 0.785

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description | Pred. No. |
|------------|-------|-------------|--------|-------|------------------------|------------------------|-----------|
| 1 | 107 | 94.7 | 993 | 11 | O35902 | DESMOGLEIN 3 (FRAGMENT | 4.01e-12 |
| 2 | 72 | 787 | 4 | OY5E8 | PROTODADHERIN BETA 15. | 1.95e-03 | |
| 3 | 65 | 57.5 | 222 | 3 | O13737 | HYPOTHETICAL 24.9 KD P | 7.00e-02 |
| 4 | 65 | 57.5 | 519 | 13 | O90425 | VENTRAL NEURAL CADHERI | 7.00e-02 |
| 5 | 63 | 55.8 | 394 | 1 | O58593 | METHYLOLOGEN-REDUCIN | 1.88e-01 |
| 6 | 63 | 55.8 | 2606 | 5 | O21920 | KL1A8.7 PROTEIN. | 1.88e-01 |
| 7 | 62 | 54.9 | 379 | 2 | O08356 | DNAP PROTEIN. | 3.06e-01 |
| 8 | 62 | 54.9 | 501 | 5 | O02213 | PUTATIVE G PROTEIN-CON | 3.06e-01 |
| 9 | 60 | 53.1 | 466 | 5 | O76322 | SYNAPSIN S-SYN-SHORT (| 7.98e-01 |
| 10 | 60 | 53.1 | 503 | 5 | O76323 | SYNAPSIN S-SYN-LONG (F | 7.98e-01 |
| 11 | 60 | 53.1 | 814 | 6 | O77704 | DESMOGLEIN TYPE 2 (FR | 7.98e-01 |
| 12 | 59 | 52.2 | 265 | 5 | O9XU08 | TO66.11 PROTEIN. | 1.28e+00 |
| 13 | 59 | 52.2 | 659 | 14 | O69468 | REPLICATION ORIGIN-BIN | 1.28e+00 |
| 14 | 59 | 52.2 | 780 | 14 | O9WT02 | ORIGIN BINDING PROTEIN | 1.28e+00 |
| 15 | 59 | 52.2 | 2163 | 5 | O01912 | SIMILARITY TO MULTIPLE | 1.28e+00 |
| 16 | 58 | 51.3 | 339 | 7 | P79570 | MHC CLASS I PRECURSOR. | 2.04e+00 |
| 17 | 58 | 51.3 | 832 | 4 | O12864 | INTESTINAL PEPTIDE-ASS | 2.04e+00 |
| 18 | 58 | 51.3 | 832 | 4 | O15336 | LI-CADHERIN. | 2.04e+00 |
| 19 | 58 | 51.3 | 5137 | 3 | O01135 | PEPTIDE SYNTHETASE. | 2.04e+00 |
| 20 | 57 | 50.4 | 235 | 2 | O67056 | DNA REPLICATION PROTEI | 3.24e+00 |

| RESULT | ID | ALIGNMENTS | PRT | 993 AA. |
|--------|--|--------------|-----|---------|
| AC | O35902 | PRELIMINARY: | | |
| DT | 01-JAN-1998 (TREMBlrel. 05, Created) | | | |
| DT | 01-JAN-1998 (TREMBlrel. 05, Last sequence update) | | | |
| DT | 01-NOV-1999 (TREMBlrel. 12, Last annotation update) | | | |
| DE | DESMOGLEIN 3 (FRAGMENT). | | | |
| GN | DSG3 | | | |
| OS | Mus musculus (Mouse). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | | |
| CC | Eutheria; Rodentia; Sciurognathi; Muridae; Mus. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-BALB/C; | | | |
| RA | ISHIKAWA H., LI K., UIRTO J. | | | |
| RL | Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases. | | | |
| CC | -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY). | | | |
| DR | EMBL; U86016; AAB65091.1; - | | | |
| DR | HSSP; P15116; INCJ. | | | |
| DR | PROSITE; PS00232; CADHERIN; 2. | | | |
| DR | PFAM; PF00028; cadherin; 4. | | | |
| KW | Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat. | | | |
| FT | NON_TER 993 | | | |
| SO | SEQUENCE 993 AA; 107888 MW; 881794BD CRC32; | | | |

| Query Match | Score 107; | DB 11; | Length 993; |
|-----------------------|--------------|---------------------|---------------------------------|
| Best Local Similarity | 86.7%; | Pred. No. 4.01e-12; | |
| Matches 13; | Conservative | 2; | Mismatches 0; Indels 0; Gaps 0; |

Db 251 CECNIKVKVDNDNFP 265

QY 1 CECNIKVKVDNDNFP 15

RESULT 2

ID OY5E8 PRELIMINARY: PRT; 787 AA.

AC OY5E8;

DT 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)

DE PROTODADHERIN BETA 15.

GN PCDH-BETA15.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria: Primates: Catarrhini; Homiidae; Homo.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 99308636.
 RA WU O., MANIATIS T.;
 RT "A striking organization of a large family of human neural cadherin-
 RT like cell adhesion genes."
 RL Cell 97:179-790(1999).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF152494; AAD43755.1; -
 DR PROSITE: PS00232; CADHERIN; 5.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SQ SEQUENCE 787 AA; 86329 MW; F66574C0 CRC32;
 Query Match 63.7%; Score 72; DB 4; Length 787;
 Best Local Similarity 66.7%; Pred. No. 1.95e-03;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Db 331 CSHSVKLVNDNFP 345
 QY 1 CECNKKVNDNFP 15

RESULT 3 PRELIMINARY: PRT: 222 AA.
 ID 013737
 AC 013737
 DT 01-JUN-1998 (TEMBLrel. 06, Created)
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
 DE HYPOTHETICAL 24.9 KD PROTEIN C16E8.02 IN CHROMOSOME 1.
 GN SPAC16E8.02.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA OLIVER K., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO YEAST YGL010W AND SOME, TO N.CRASSA ATP-6.
 DR EMBL: Z85529; CAB1031.1; -
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 27 47 POTENTIAL.
 FT TRANSMEM 65 85 POTENTIAL.
 FT TRANSMEM 105 125 POTENTIAL.
 FT TRANSMEM 135 155 POTENTIAL.
 FT DOMAIN 30 33 POLY-LEU.
 SQ SEQUENCE 222 AA; 24993 MW; 396C437E CRC32;
 Query Match 57.5%; Score 65; DB 3; Length 222;
 Best Local Similarity 66.7%; Pred. No. 7.00e-02;
 Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Db 167 NIKLKGVNDNFP 178
 QY 4 NIKVKVNDNFP 15
 RESULT 4 PRELIMINARY: PRT: 519 AA.
 ID 090425
 AC 090425
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
 DE VENTRAL NEURAL CADHERIN (FRAGMENT).
 GN VNC.
 OS Brachydanio rerio (zebrafish) (zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
 OC Cyprinoidae; Cyprinidae; Rasbortinae; Danio.
 RN (1)

RP SEQUENCE FROM N.A.
 RA FRANKLIN J.L., SARGENT T.D.;
 RC Dev. Dyn. 206:0-0(0).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: U41419; AAB47406.1; -
 DR HSSP: P15116; INCJ.
 DR ZFIN: ZDB-GENE-980526-170; vnc.
 DR PROSITE: PS00232; CADHERIN; 1.
 DR PFAM: PF00028; cadherin; 3.
 DR PFAM: PF01049; cadherin_C-term; 1.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 FT NON_TER 1
 SQ SEQUENCE 519 AA; 57807 MW; E6DA0079 CRC32;
 Query Match 57.5%; Score 65; DB 13; Length 519;
 Best Local Similarity 91.7%; Pred. No. 7.00e-02;
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 202 NIKVKVNDNFP 213
 QY 4 NIKVKVNDNFP 15

RESULT 5 PRELIMINARY: PRT: 394 AA.
 ID 058593
 AC 058593
 DT 01-AUG-1998 (TEMBLrel. 07, Created)
 DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
 DE METHYLVIOLIGEN-REDUCING HYDROGENASE POLYFERREDOXIN PROTEIN.
 GN VHUB OR M1193.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 2661;
 RX MEDLINE: 96337999.
 RA BULT C.G., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOGAYNE J.D.,
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
 RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUJIMANN J.L., NGUYEN D.,
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
 RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii."
 RL Science 273:1058-1073(1996).
 CC -1- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
 CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
 CC -1- COPFACTOR: BINDS TEN 4FE-4S CLUSTER.
 CC -1- SUBUNIT: VHUB CONSISTS OF THE VHUA, VHUG, VHUU SUBUNTS AND A
 CC FERREDOXIN PROTEIN.
 CC -1- MISCELLANEOUS: THIS PROTEIN CONTAINS SEVERAL REPEATED FERREDOXIN-
 CC LIKE DOMAINS.
 DR EMBL: U67560; AAB99195.1; -
 DR HSSP: P00185; ICLF.
 DR TIGR: M01193; -
 DR PFAM: PF00037; fer4; 6.
 DR PRINTS: PR00353; 4FE4SFROXIN.
 KW Electron transport; Iron-sulfur; Repeat.
 FT METAL 13 13
 FT METAL 16 16 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 19 19 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 23 23 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 26 26 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 30 30 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 33 33 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 36 36 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 39 39 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 42 42 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 45 45 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 48 48 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 51 51 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 54 54 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 57 57 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).

```

FT METAL 79 79 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 83 83 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 101 101 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 104 104 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 107 107 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 111 111 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 142 142 IRON-SULFUR 5 (4FE-4S) (BY SIMILARITY).
FT METAL 145 145 IRON-SULFUR 5 (4FE-4S) (BY SIMILARITY).
FT METAL 148 148 IRON-SULFUR 5 (4FE-4S) (BY SIMILARITY).
FT METAL 152 152 IRON-SULFUR 5 (4FE-4S) (BY SIMILARITY).
FT METAL 171 171 IRON-SULFUR 6 (4FE-4S) (BY SIMILARITY).
FT METAL 174 174 IRON-SULFUR 6 (4FE-4S) (BY SIMILARITY).
FT METAL 177 177 IRON-SULFUR 6 (4FE-4S) (BY SIMILARITY).
FT METAL 181 181 IRON-SULFUR 6 (4FE-4S) (BY SIMILARITY).
FT METAL 210 210 IRON-SULFUR 7 (4FE-4S) (BY SIMILARITY).
FT METAL 213 213 IRON-SULFUR 7 (4FE-4S) (BY SIMILARITY).
FT METAL 216 216 IRON-SULFUR 7 (4FE-4S) (BY SIMILARITY).
FT METAL 220 220 IRON-SULFUR 7 (4FE-4S) (BY SIMILARITY).
FT METAL 240 240 IRON-SULFUR 8 (4FE-4S) (BY SIMILARITY).
FT METAL 243 243 IRON-SULFUR 8 (4FE-4S) (BY SIMILARITY).
FT METAL 246 246 IRON-SULFUR 8 (4FE-4S) (BY SIMILARITY).
FT METAL 250 250 IRON-SULFUR 8 (4FE-4S) (BY SIMILARITY).
FT METAL 306 306 IRON-SULFUR 9 (4FE-4S) (BY SIMILARITY).
FT METAL 309 309 IRON-SULFUR 9 (4FE-4S) (BY SIMILARITY).
FT METAL 312 312 IRON-SULFUR 9 (4FE-4S) (BY SIMILARITY).
FT METAL 316 316 IRON-SULFUR 9 (4FE-4S) (BY SIMILARITY).
FT METAL 343 343 IRON-SULFUR 10 (4FE-4S) (BY SIMILARITY).
FT METAL 346 346 IRON-SULFUR 10 (4FE-4S) (BY SIMILARITY).
FT METAL 349 349 IRON-SULFUR 10 (4FE-4S) (BY SIMILARITY).
FT METAL 353 353 IRON-SULFUR 10 (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 394 AA: 43342 MW: 9C45BC4F CRC32:

```

```

Query Match 55.8% Score 63: DB 1: Length 394:
Best Local Similarity 53.8% Pred. No. 1.88e-01:
Matches 7: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

```

```

Db 83 CPICIKKVDNDF 95
Oy 3 CNIKKVDNDF 15

```

```

RESULT 6 PRELIMINARY: PRT: 2606 AA.
ID Q21920: Q21927:
AC Q21920: Q21927:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE R11A8.7 PROTEIN.
GN R11A8.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nemata; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA BARBIL S.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIR M., JOHNSTON L.,
RA JONES M., KERSHAM J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALLON N., SMITH A., SONNHAMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [3]

```

```

RP SEQUENCE FROM N.A.
RA COMINGS P.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: 270686; CA94615.1; -.
DR EMBL: 270310; CA94615.1; JOINED.
DR EMBL: 270310; CA94370.1; -.
DR EMBL: 270686; CA94370.1; JOINED.
DR HSSP: Q00421; IAWC.
DR PFW: PFW0023; ank; 19.
DR PFW: PFW0013; KH-domain; 1.
SQ SEQUENCE 2606 AA: 285415 MW: F4D767A8 CRC32:

```

```

Query Match 55.8% Score 63: DB 5: Length 2606:
Best Local Similarity 53.8% Pred. No. 1.88e-01:
Matches 7: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

```

```

Db 965 CAISVDMESNFP 977
Oy 3 CNIKKVDNDF 15

```

```

RESULT 7 PRELIMINARY: PRT: 379 AA.
ID Q08356: Q08356:
AC Q08356: Q08356:
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE DNAS PROTEIN.
GN DNAS.
OS Rhodospseudomonas sp.
OC Bacteria; Proteobacteria; alpha subdivision; Bradyrhizobium group;
OC Rhodospseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-NO. 7;
RX MEDLINE: 97214635.
RA MOMMA K., INUI M., YAMAGATA H., YUKAWA H.;
RT "Cloning of dnaK and dnaJ homologous genes from a purple non-sulfur
RT bacterium Rhodospseudomonas species."
RL Biochim. Biophys. Acta 150:235-239(1997).
CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAS, AND TO EUKARYOTIC DNAS-LIKE
CC PROTEINS.
DR EMBL: D78133; BAA19797.1; -.
DR HSSP: P08622; 1XBL.
DR PROSITE: PS00636; DNAS_1; 1.
DR PROSITE: PS00637; DNAS_CXXCXGXG; 1.
DR PFW: PFW00226; Dnas; 1.
DR PFW: PFW0684; Dnas_CXXCXGXG; 1.
DR PFW: PFW1556; Dnas-C; 1.
DR PRINTS: PRO0625; DNASPROTEIN.
KW Chapterone: DNA replication.
SQ SEQUENCE 379 AA: 40992 MW: 5D28C6EE CRC32:

```

```

Query Match 54.9% Score 62: DB 2: Length 379:
Best Local Similarity 38.5% Pred. No. 3.06e-01:
Matches 5: Conservative 6: Mismatches 2: Indels 0: Gaps 0:

```

```

Db 44 QCEIKKVDNDF 56
Oy 2 CNIKKVDNDF 14

```

```

RESULT 8 PRELIMINARY: PRT: 501 AA.
ID Q02213: Q02213:
AC Q02213: Q02213:
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE PUTATIVE G PROTEIN-COUPLED RECEPTOR C02D4.2.
GN C02D4.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nemata; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA BAYNES C.;
RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PUTATIVE G-PROTEIN RECEPTOR.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC MOST SIMILAR TO INSECT OCTOPAMINE RECEPTORS.
DR EMBL: 281031; CAB02718.1; -.
DR WORMPEP: C02D4.2; CE07837.
DR PROSITE: PS00237; G_PROTEIN_RECEPTOR. 1.
DR PFM: PF00001; 7tm1.1; 2.
DR PRINTS: PR00237; GPCRHHODPSN.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 49
FT TRANSMEM 50 72
FT DOMAIN 73 82
FT TRANSMEM 83 104
FT TRANSMEM 120 136
FT TRANSMEM 137 157
FT TRANSMEM 158 177
FT TRANSMEM 178 200
FT TRANSMEM 201 224
FT TRANSMEM 225 246
FT TRANSMEM 247 373
FT TRANSMEM 374 395
FT TRANSMEM 396 426
FT TRANSMEM 427 447
FT TRANSMEM 448 487
FT DOMAIN 493 496
FT CARBOHYD 15 15
FT DISULFID 134 212
FT SEQUENCE 501 AA; 56404 MW; 18F1C57F CRC32;
BY SIMILARITY.

Query Match 54.9%; Score 62; DB 5; Length 501;
Best Local Similarity 54.5%; Pred. No. 3,06e-01;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DB 336 CKYKVDVND 346
OY 3 CNIKYKVDVND 13

RESULT 9
ID 076322 PRELIMINARY; PRT; 466 AA.
AC 076322:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE SYNAPSIN S-SYN-SHORT (FRAGMENT).
OS Loligo pealeii (Longfin squid).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Teuthoidea;
OC Myopsida; Loliginidae; Loligo.
RN [1]
RP SEQUENCE FROM N.A.
RA HILFiker S., SCHWEITZER F.E., KAO H.-T., CZERNIK A.J., GREENGARD P.,
RA AUGUSTINE G.J.;
RT "Two Sites of Action for Synapsin Domain E in Regulating
RT Neurotransmitter Release."
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF055373; AAC24822.1; -.
DR HSSP: P17599; IAUx.
FT NON_TER 1
FT SEQUENCE 466 AA; 52133 MW; C751B337 CRC32;

Query Match 53.1%; Score 60; DB 5; Length 466;
Best Local Similarity 35.7%; Pred. No. 7,98e-01;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

DB 287 CKCDIHYOKIGNNY 300
OY 1 CECNIKVDVNDNF 14
```

```

RESULT 10
ID 076323 PRELIMINARY; PRT; 503 AA.
AC 076323:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE SYNAPSIN S-SYN-LONG (FRAGMENT).
OS Loligo pealeii (Longfin squid).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Teuthoidea;
OC Myopsida; Loliginidae; Loligo.
RN [1]
RP SEQUENCE FROM N.A.
RA HILFiker S., SCHWEITZER F.E., KAO H.-T., CZERNIK A.J., GREENGARD P.,
RA AUGUSTINE G.J.;
RT "Two Sites of Action for Synapsin Domain E in Regulating
RT Neurotransmitter Release."
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF055374; AAC24823.1; -.
DR HSSP: P17599; IAUx.
FT NON_TER 1
FT SEQUENCE 503 AA; 55738 MW; 693971D9 CRC32;

Query Match 53.1%; Score 60; DB 5; Length 503;
Best Local Similarity 35.7%; Pred. No. 7,98e-01;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

DB 287 CKCDIHYOKIGNNY 300
OY 1 CECNIKVDVNDNF 14

RESULT 11
ID 077704 PRELIMINARY; PRT; 814 AA.
AC 077704:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE DESMOCOLLIN TYPE 2 (FRAGMENT).
GN DSC2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 98380123.
RA ROBERTS G.A., BURDETT I.D., PIDSLEY S.C., KING I.A., MAGEE A.I.,
RA BUXTON R.S.;
RT "Antisense expression of a desmocollin gene in MDCK cells alters
RT desmosome plaque assembly but does not affect desmoglein expression."
RL Eur. J. Cell Biol. 76:192-203(1998).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AJ002299; CAA05309.1; -.
DR HSSP: P09803; ISUH.
DR PROSITE: PS00232; CADHERIN; 3.
DR PFM: PF00028; cadherin; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON_TER 1
FT CHAIN 101 814
FT SEQUENCE 814 AA; 91081 MW; BE0007A8 CRC32;

Query Match 53.1%; Score 60; DB 6; Length 814;
Best Local Similarity 61.5%; Pred. No. 7,98e-01;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 306 CIINIDVNDLP 318
OY 3 CNIKYKVDVNDNF 15

RESULT 12
ID 09XU08 PRELIMINARY; PRT; 265 AA.
AC 09XU08:
```

DT 01-NOV-1999 (TEMBLrel. 12, Created)
 DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
 DE T06G6.11 PROTEIN.
 GN T06G6.11.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA KERSHAM J.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans." (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAM J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL: 281587; CAB04709.1;
 SO SEQUENCE 265 AA; 30553 MW; 8CC2E535 CRC32;

Query Match 52.2%; Score 59; DB 5; Length 265;
 Best Local Similarity 46.2%; Pred. No. 1.28e+00;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 Db 190 DCNVRMKNDPNF 202
 Oy 2 ECNKKVQVNDNF 14

RESULT 13
 ID 069468; PRELIMINARY; PRT; 659 AA.
 AC 069468;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-NOV-1996 (TEMBLrel. 08, Last annotation update)
 DE REPLICATION ORIGIN-BINDING PROTEIN (FRAGMENT).
 GN HDRFO.
 OS Human herpesvirus 6.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Roseolovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-U1102;
 RX MEDLINE: 95027704.
 RA NICHOLAS J.;
 RT "Nucleotide sequence analysis of a 21-kbp region of the genome of
 human herpesvirus-6 containing homologues of human cytomegalovirus
 RT major immediate-early and replication genes.";
 RL Virology 204:738-750(1994).
 DR EMBL: U13194; AAA68464.1;
 FT NON_TER 1
 SO SEQUENCE 659 AA; 75779 MW; 65BDEAB0 CRC32;

Query Match 52.2%; Score 59; DB 14; Length 659;
 Best Local Similarity 53.3%; Pred. No. 1.28e+00;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 Db 303 CENSIRVDVNGNFP 317
 Oy 11 : 1111 1 : 11

Oy 1 CECNKKVQVNDNF 15
 RESULT 14
 ID 09WT02; PRELIMINARY; PRT; 780 AA.
 AC 09WT02;
 DT 01-NOV-1999 (TEMBLrel. 12, Created)
 DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
 DE ORIGIN BINDING PROTEIN.
 GN U73.
 OS Human herpesvirus 6.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Roseolovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HST;
 RA ISEGAWA Y., MUKAI T., NAKANO K., KAGAWA M., CHEN J., MORI Y.,
 RA SUNAGAWA T., SASHIHARA J., ZOU P., KOSUGE H., YAMANISHI K.;
 RT "A comparison of the complete DNA sequences between human herpesvirus-
 RT 6 variant A and B.";
 RL J. Virol. 0:0-0(1999).
 DR EMBL: AB021506; BAA78294.1;
 SO SEQUENCE 780 AA; 89540 MW; 5F3A240A CRC32;

Query Match 52.2%; Score 59; DB 14; Length 780;
 Best Local Similarity 53.3%; Pred. No. 1.28e+00;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 Db 424 CENSIRVDVNGNFP 438
 Oy 1 CECNKKVQVNDNF 15

RESULT 15
 ID 001912; PRELIMINARY; PRT; 2163 AA.
 AC 001912;
 DT 01-JUL-1997 (TEMBLrel. 04, Created)
 DT 01-JUL-1997 (TEMBLrel. 04, Last sequence update)
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
 DE SIMILARITY TO MULTIPLE CADHERIN-TYPE REPEATS.
 GN R10F2.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAM J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 SO SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2;
 RA DU Z., GATTUNG S.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATSON R.;
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

DR EMBL: AF003388; AAB54266.1; -
DR HSSP: P15116; INCI.
DR PROSITE: PS00232; CADHERIN; 8.
DR PFAM: PF00001; 7tm_1; 1.
DR PFAM: PF00028; cadherin; 15.
DR PRINTS: PR00205; CADHERIN.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ SEQUENCE 2165 AA: 238609 MW: 6085C652 CRC32;

| | | | | |
|-----------------------|-------------------|--------------|----------|--------------|
| Query Match | 52.2% | Score 59 | DB 5 | Length 2163 |
| Best Local Similarity | 53.3% | Pred. NO. | 1.28e+00 | |
| Matches | 8 | Conservative | 3 | Mismatches 4 |
| | | | | Indels 0 |
| | | | | Gaps 0 |
| Db | 1629 CKCHVIVDNDNP | 1643 | | |
| | | | | |
| Oy | 1 CECNIRKVDVNDNP | 15 | | |

Search completed: Sat May 13 07:21:16 2000
Job time : 154 secs.

THIS PAGE BLANK (USPTO)

 W0846 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 07:29:32 2000; Mpsrch time 3.02 Seconds
 Tabular output not generated. 117.801 Million cell updates/sec

Title: >US-08-991-628-6
 Description: (1-15) from US08991628.pap
 Perfect Score: 109
 Sequence: 1 SARTLNRRYTGPTTF 15

Scoring table:
 Gap 15
 PAM 150

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-geneseq35
 1:geneseqp

Statistics: Mean 19.232; Variance 58.683; scale 0.328

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|------------|-------|-------------|--------|-------|-------------------------|-----------|
| 1 | 109 | 100.0 | 15 | 1 | Self epitope of desmog | 5.72e-05 |
| 2 | 109 | 100.0 | 614 | 1 | Pemphigus vulgaris ant | 5.72e-05 |
| 3 | 109 | 100.0 | 999 | 1 | Human pemphigus vulgar | 5.72e-05 |
| 4 | 56 | 51.4 | 480 | 1 | Haem 84-2 portion of H | 6.87e+01 |
| 5 | 56 | 51.4 | 516 | 1 | Haem 84-1 portion of H | 6.87e+01 |
| 6 | 55 | 50.5 | 431 | 1 | Mouse CRF RBL receptor | 8.71e+01 |
| 7 | 54 | 49.5 | 91 | 1 | N-terminal region of b | 1.10e+02 |
| 8 | 54 | 49.5 | 246 | 1 | Protein able to bind t | 1.10e+02 |
| 9 | 54 | 49.5 | 428 | 1 | Chimeric KMWLV and Ty3 | 1.10e+02 |
| 10 | 54 | 49.5 | 429 | 1 | Chimeric KMWLV and Ty3 | 1.10e+02 |
| 11 | 54 | 49.5 | 431 | 1 | Rat CRF-beta receptor | 1.10e+02 |
| 12 | 54 | 49.5 | 449 | 1 | Chimeric KMWLV and Ty3 | 1.10e+02 |
| 13 | 54 | 49.5 | 1079 | 1 | MULV reverse transcript | 1.10e+02 |
| 14 | 54 | 49.5 | 1196 | 1 | Osteoinductive retrovi | 1.10e+02 |
| 15 | 53 | 48.6 | 105 | 1 | Variable light chain o | 1.40e+02 |
| 16 | 53 | 48.6 | 105 | 1 | Anti-human FasL antilo | 1.40e+02 |
| 17 | 53 | 48.6 | 653 | 1 | IDUA. | 1.40e+02 |
| 18 | 53 | 48.6 | 2115 | 1 | Rubella virus RA27/3 N | 1.40e+02 |
| 19 | 53 | 48.6 | 2205 | 1 | Infectious rubella vir | 1.40e+02 |
| 20 | 52 | 47.7 | 105 | 1 | Antibody Lp1/2-6-3-VL | 1.77e+02 |
| 21 | 52 | 47.7 | 105 | 1 | Antibody Lp1/2-6-3-VL | 1.77e+02 |
| 22 | 52 | 47.7 | 140 | 1 | Human 5' EST secreted | 1.77e+02 |
| 23 | 52 | 47.7 | 454 | 1 | Kaposi's sarcoma assoc | 1.77e+02 |

| | | | | | | | |
|----|----|------|------|---|--------|-------------------------|----------|
| 24 | 52 | 47.7 | 454 | 1 | R97842 | Kaposi's sarcoma assoc | 1.77e+02 |
| 25 | 52 | 47.7 | 575 | 1 | Y00157 | Enterococcus faecalis | 1.77e+02 |
| 26 | 52 | 47.7 | 601 | 1 | Y00156 | Enterococcus faecalis | 1.77e+02 |
| 27 | 52 | 47.7 | 691 | 1 | R04711 | Sequence of guinea pig | 1.77e+02 |
| 28 | 52 | 47.7 | 932 | 1 | W97701 | Staphylococcus aureus | 1.77e+02 |
| 29 | 52 | 47.7 | 1844 | 1 | W18302 | Photobacterium luminesc | 1.77e+02 |
| 30 | 52 | 47.7 | 1844 | 1 | W56558 | Toxin TcdA, encoded by | 1.77e+02 |
| 31 | 52 | 47.7 | 2504 | 1 | W56557 | Toxin TcdA, encoded by | 1.77e+02 |
| 32 | 52 | 47.7 | 2504 | 1 | W1871 | Photobacterium luminesc | 1.77e+02 |
| 33 | 51 | 46.8 | 39 | 1 | R25101 | bGRF protidg analogue | 2.23e+02 |
| 34 | 51 | 46.8 | 106 | 1 | Y01034 | Anti VLA-4 antibody HP | 2.23e+02 |
| 35 | 51 | 46.8 | 106 | 1 | W66742 | Light chain variable r | 2.23e+02 |
| 36 | 51 | 46.8 | 127 | 1 | W06443 | Humc3 VL region. | 2.23e+02 |
| 37 | 51 | 46.8 | 128 | 1 | W72435 | Humanised VK sequence. | 2.23e+02 |
| 38 | 51 | 46.8 | 128 | 1 | W72433 | Transplanted VK sequen | 2.23e+02 |
| 39 | 51 | 46.8 | 128 | 1 | R55211 | Humanised HP1/2 light | 2.23e+02 |
| 40 | 51 | 46.8 | 128 | 1 | Y01036 | Anti VLA-4 antibody SV | 2.23e+02 |
| 41 | 51 | 46.8 | 128 | 1 | R55209 | Humanised HP1/2 kappa | 2.23e+02 |
| 42 | 51 | 46.8 | 128 | 1 | R59935 | VK3 (DOWDY) VL. | 2.23e+02 |
| 43 | 51 | 46.8 | 845 | 1 | W56579 | Fragment of toxin TcdA | 2.23e+02 |
| 44 | 51 | 46.8 | 1261 | 1 | W93601 | Tomato Xa21 clone TRK2 | 2.23e+02 |
| 45 | 51 | 46.8 | 2516 | 1 | W56572 | Toxin TcdA, encoded by | 2.23e+02 |

ALIGNMENTS

RESULT 1
 ID W0846 standard; peptide: 15 AA.
 AC W04846;
 DT 18-FEB-1997 (first entry)
 DE Self epitope of desmoglein 3, implicated in autoimmune disease.
 KW Tolerisation: self-epitope; antigen: autoimmune disease.
 KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
 KW pemphigus vulgaris; desmoglein; multiple sclerosis;
 KW herpes simplex virus; adenovirus; phosphomannomutase;
 KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
 KW influenza; haemagglutinin; reovirus; sigma protein.
 OS Homo sapiens.
 ON W09627387-A1.
 ND 12-SEP-1996.
 PD 07-MAR-1996; US03182.
 PR 07-MAR-1995; US-400796.
 PA (HARD) HARVARD COLLEGE.
 PI Strominger JL, Wucherpfennig KW;
 DR WPT. 96-425218/42.
 PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
 PT antigens - useful in disease treatment, and method for
 PT identification of other self and non-self antigens implicated in
 PT auto-immune disease
 PS Claim 1: Page 41: 58pp: English.
 CC Pharmaceutical preparations for tolerisation to antigens comprise
 CC either an isolated human non-collagen or non-myosin basic protein
 CC (MBP) polypeptide which is capable of tolerising an individual to an
 CC autoantigen; or an isolated human pathogen polypeptide capable of
 CC tolerising an individual to that polypeptide. In both cases, the
 CC polypeptide (whether self or non-self) includes an amino acid
 CC sequence corresponding to a sequence motif for a MHC class II
 CC protein, such as HLA-DR, which is associated with a human autoimmune
 CC disease and which binds to the polypeptide to activate autoreactive
 CC T-cells in individuals with the autoimmune disease. This peptide is
 CC derived from the human desmoglein 3 protein (amino acids 512-526)
 CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
 CC derived from the human desmoglein protein are described in W04841-47.
 SO Sequence 15 AA:

Query Match 100.0%; Score 109; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 5.72e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 SARTLNRRYTGPTTF 15
 Oy 1 SARTLNRRYTGPTTF 15

RESULT 2
ID W07908 standard; protein: 614 AA.
AC W07908:
DT 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
dermatology.
OS Homo sapiens.
PN 308188540-A.
PD 23-JUL-1996.
PF 30-JUN-1995; 165632.
PR 30-JUN-1994; JP-173291.
PA (NISH/) NISHUKAWA T.
DR WPI: 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris autoantibody -
useful to treat and diagnose pemphigus vulgaris
PS Claim 1: Page 7-9; 9pp; Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
extracellular region. The PV antigen is produced in patients with
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose
CC pemphigus vulgaris.
SQ Sequence 614 AA:

Query Match 100.0%; Score 109; DB 1; Length 614;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 511 SARTLNRRYGPYTF 525
QY 1 SARTLNRRYGPYTF 15

RESULT 3
ID R30742 standard; Protein: 999 AA.
AC R30742:
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130kD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US7798918-A.
PD 15-DEC-1992.
PF 27-NOV-1991; 798918.
PR 27-NOV-1991; US-798918.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagai M, Klaus-Kovtun V, Stanley JR;
DR WPI: 93-067436/08.
N-PSDB: Q35992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses
PS Disclosure; Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA:

Query Match 100.0%; Score 109; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 512 SARTLNRRYGPYTF 526
QY 1 SARTLNRRYGPYTF 15

RESULT 4

ID R12099 standard; Protein: 480 AA.
AC R12099:
DT 22-JUL-1991 (first entry)
DE Haem 84-2 portion of H. contortus 55A surface protein.
KW Parasitic nematode; vaccine; Hc.
OS Haemonchus contortus.
PN AU9062569-A.
PD 21-MAR-1991.
PF 17-SEP-1990; 062569.
PR 18-SEP-1989; US-408339.
PR 01-MAR-1990; US-487181.
PA (SYNE-) SYNERGEN INC.
DR WPI: 91-133285/19.
DR N-PSDB: Q11799.
PT Protein from Haemonchus contortus and other nematodes - used as
PT therapeutic and prophylactic agent to protect plants, animals or
PT humans from parasitic nematode infection.
PS Disclosure; Fig 49; 209pp; English.
CC The proteins derived from the nematode DNA may be used to derive
CC vaccines against parasitic infection of plants, humans and animals
CC esp. sheep. Mabs may also be raised to provide passive therapy and
CC prophylaxis against infection.
SQ Sequence 480 AA:

Query Match 51.4%; Score 56; DB 1; Length 480;
Best Local Similarity 54.5%; Pred. No. 6.87e-01;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DB 376 VNNRYGGRYK 386
QY 5 LNNRYGPYTF 15

RESULT 5
ID R12098 standard; Protein: 516 AA.
AC R12098:
DT 22-JUL-1991 (first entry)
DE Haem 84-1 portion of H. contortus 55A surface protein.
KW Parasitic nematode; vaccine; Hc.
OS Haemonchus contortus.
PN AU9062569-A.
PD 21-MAR-1991.
PF 17-SEP-1990; 062569.
PR 18-SEP-1989; US-408339.
PR 01-MAR-1990; US-487181.
PA (SYNE-) SYNERGEN INC.
DR WPI: 91-133285/19.
DR N-PSDB: Q11798.
PT Protein from Haemonchus contortus and other nematodes - used as
PT therapeutic and prophylactic agent to protect plants, animals or
PT humans from parasitic nematode infection.
PS Disclosure; Fig 49; 209pp; English.
CC The proteins derived from the nematode DNA may be used to derive
CC vaccines against parasitic infection of plants, humans and animals
CC esp. sheep. Mabs may also be raised to provide passive therapy and
CC prophylaxis against infection.
SQ Sequence 516 AA:

Query Match 51.4%; Score 56; DB 1; Length 516;
Best Local Similarity 54.5%; Pred. No. 6.87e-01;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DB 484 VNNRYGGRYK 494
QY 5 LNNRYGPYTF 15

RESULT 6
ID R97293 standard; Protein: 431 AA.
AC R97293:
DT 21-AUG-1996 (first entry)
DE Mouse CRF RBL receptor.
KW Corticotropin releasing factor receptor; CRF-R; corticoidiberlin;

```

KW signal transduction.
OS Mus sp.
FH key Location/Qualifiers
FT modified_site 52
FT modified_site 61 /label= N-glycosylation_site
FT modified_site 61 /label= N-glycosylation_site
FT modified_site 94 /label= N-glycosylation_site
FT modified_site 105 /label= N-glycosylation_site
FT modified_site 113 /label= N-glycosylation_site
FT domain 139.1159 /label= N-glycosylation_site
FT domain 169.1188 /label= Transmembrane_domain-1
FT domain 169.1188 /label= Transmembrane_domain-2
FT domain 206.1229 /label= Transmembrane_domain-3
FT domain 245.1265 /label= Transmembrane_domain-4
FT domain 285.1307 /label= Transmembrane_domain-5
FT domain 331.1351 /label= Transmembrane_domain-6
FT domain 365.1385 /label= Transmembrane_domain-7
FT domain 365.1385 /label= Transmembrane_domain-7
PN WO9617934-A2.
PD 13-JUN-1996.
PF 06-DEC-1995: U15909.
PR 09-DEC-1994: US-353537.
PR 17-JAN-1995: US-374009.
PA (SALK ) SALK INST BIOLOGICAL STUDIES.
PI Chen R, Donaldson CJ, Lewis KA, Perrin MH, Sawchenko P;
PI Vale WJ; 287179/29.
DR N-PSDB: T28972.
PT Isolated corticotropin-releasing factor receptor (CRF-R) - used to
PT develop prods. for modulating signal transduction activity mediated
PT by CRF-R.
PS Claim 3; Page 83-85; 102pp; English.
CC Mouse corticotropin releasing factor receptor mCRF-R1 was
CC identified as the product of a cDNA clone (T28972) isolated from a
CC mouse heart library. Recombinant mCRF-R1 can be expressed in
CC host cells transformed by the cDNA clone. The receptor can be used
CC to identify agonists and antagonists that modulate the signal
CC transduction activity mediated by CRF receptors. It may be
CC administered therapeutically to reduce high ACTH levels caused by
CC excess CRF.
SQ Sequence 431 AA;

Query Match 50.5%; Score 55; DB 1; Length 431;
Best Local Similarity 46.2%; Pred. No. 8.71e+01;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 47 RTTGNFSGPYTY 59
   |||:|||||:
OY 3 RTLNRYTGPYTF 15

RESULT 7
ID P61040 standard; protein: 91 AA.
AC P61040:
DE 03-JUL-1991 (first entry)
DE N-terminal region of bacterial expression vector.
KW Galactose: expression vector.
OS Synthetic.
PN J60248181-A.
PD 07-DEC-1985.
PF 23-MAY-1984: 102685.
PR 23-MAY-1984: JP-102685.
PA (SHTS ) SHITSIDO KK.
DR WPI: 86-025462/04.

```

```

DR N-PSDB: N60836.
PT Enzyme expression vector - giving transformant enzymes which can
PT be used as host enzymes in polypeptide biosynthesis.
PS Disclosure; Fig 4; 12pp; Japanese.
CC The vector may be used to express heterologous genes from a
CC transformed host when galactose is added as a carbon source.
SQ Sequence 91 AA;

Query Match 49.5%; Score 54; DB 1; Length 91;
Best Local Similarity 54.5%; Pred. No. 1.10e+02;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 67 LNPRESYTYF 77
   |||:|||||:
OY 5 LNNRYTGPYTF 15

RESULT 8
ID R40924 standard; protein: 246 AA.
AC R40924:
DE 17-FEB-1994 (first entry)
DE Protein able to bind to HIV-1 tat protein.
KW Antibodies; antigen binding proteins; library; HIV;
KW Human Immunodeficiency Virus.
OS Synthetic.
PN EP-557897-A.
PD 01-SEP-1993.
PF 19-FEB-1993: 102609.
PR 28-FEB-1992: US-843125.
PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.
PI Dillon PJ, Rosen CA;
PI WPI: 93-274375/35.
DR N-PSDB: Q48605.
PT Antigen-binding proteins and corresp. synthetic genes - are
PT constructed synthetically and used to make antibody library, in
PT disease diagnosis, etc.
PS Claim 13; Page 22; 40pp; English.
CC The synthetic gene encodes a protein corresponding to an antibody
CC capable of binding to a specific antigen. In this case the HIV-1 tat
CC protein. Many synthetic genes are synthesised, each containing a
CC predetermined nucleotide region encoding the framework nucleotide
CC the heavy and light chains of antibody and underdetermined nucleotide
CC regions which are random sequences. The genes are then used in the
CC construction of vectors which are subsequently used to transform
CC microbes. The proteins thus produced are screened for binding
CC activity to the specific antigen.
SQ Sequence 246 AA;

Query Match 49.5%; Score 54; DB 1; Length 246;
Best Local Similarity 42.9%; Pred. No. 1.10e+02;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 57 ARYLDGSGFNGRYTI 70
   |||:|||||:
OY 2 ARTLNRYTGPYTF 15

RESULT 9
ID W08609 standard; protein: 428 AA.
AC W08609:
DE 16-JUL-1997 (first entry)
DE Chimeric MoMuV and Ty3 integrase designated Atbmcm.
KW Moloney murine leukemia virus; saccharomyces cerevisiae;
KW retrotransposon; yeast Ty3; position specific integration;
KW inflammation; cardiovascular; autoimmune; cancer; HIV; haemophilia;
KW Alzheimer's disease; rheumatoid arthritis; chimeric.
OS Chimeric - Moloney murine leukemia virus.
KW Chimeric - Saccharomyces cerevisiae transposon Ty3.
FH key location/Qualifiers
FT domain 1.61
FT /label= A
FT /note= "From Ty3"
FT domain 62.283

```

```

FT      /label-B
PI      /note-"From MOMLV"
PI      284. .428
FT      /label-C
FT      /note-"From MOMLV"
PI      W0637626-A1.
PD      28-NOV-1996.
PF      10-MAY-1996: U06727.
PR      22-MAY-1995: US-445466.
PA      (CHIR ) CHIRON VIAGENE INC.
PI      (REGG ) UNIV CALIFORNIA.
PI      Billachone VM, Dillidine SL, Jolly DJ, Respass JG;
PI      Sandmeyer SB;
PI      WPI: 97-021229/02.
PT      New chimeric integrase for targeted vector integration in
PT      eukaryotic genomes - useful for gene therapy, providing more
PT      consistent gene expression and lower rates of insertional
PT      mutagenesis
PS      Claim 7: Page -: 98pp; English.
CC      The present sequence is a specific example of a chimeric integrase
CC      made up of three domains from MOMLV and Ty3, with at least one domain
CC      derived from Ty3 integrase. This protein can direct integration of a
CC      vector construct into a defined region of a target eukaryotic genome.
CC      As part of gene delivery vehicles and transduction competent
CC      recombinant retroviral particles it can be useful in somatic
CC      and germ cell gene therapy (in vivo or ex vivo) of a wide range of
CC      genetic, infectious, degenerative, inflammatory, cardiovascular and
CC      autoimmune diseases or cancer. Typical examples of the many potential
CC      applications include treatment of HIV infection, haemophilia, Alzheimer's
CC      disease and rheumatoid arthritis. This protein rather than wild-type
CC      retroviral integrase reduces the rate of insertional mutagenesis and the
CC      degree of variation in gene expression, and particularly does not disrupt
CC      tRNA gene expression.
CC      N.B. The present sequence is not shown in the specification, but is
CC      made up from the two original sequences as specified, MOMLV and Ty3,
CC      which are shown.
SQ      Sequence 428 AA:

Query Match      49.5%; Score 54; DB 1; Length 428;
Best Local Similarity 46.2%; Pred. No. 1.10e+02;
Matches      6; Conservative      3; Mismatches      4; Indels      0; Gaps      0;

DB      360 TKNLEPRMKGPGYT 372
      : : : : :
OY      2 ARTLNRYTGPYT 14

RESULT 10
ID      W08606 standard; protein: 429 AA.
AC      W08606;
DT      16-JUL-1997 (first entry)
DE      Chimeric MOMLV and Ty3 integrase designated Ambtcm.
KW      Moloney murine leukaemia virus; Saccharomyces cerevisiae;
KW      retrotransposon; yeast Ty3; position specific integration;
KW      inflammation; cardiovascular; autoimmune; cancer; HIV; haemophilia;
KW      Alzheimer's disease; rheumatoid arthritis; chimeric.
OS      Chimeric - Moloney murine leukaemia virus.
OS      Chimeric - Saccharomyces cerevisiae transposon Ty3.
FH      key
      location/Qualifiers
FT      domain
FT      1. .41
FT      /label-A
FT      /note-"From MOMLV"
FT      42. .284
FT      /label-B
FT      /note-"From Ty3"
FT      285. 429
FT      /label-C
FT      /note-"From MOMLV"
PI      W0637626-A1.
PD      28-NOV-1996.
PF      10-MAY-1996: U06727.
PR      22-MAY-1995: US-445466.
PA      (CHIR ) CHIRON VIAGENE INC.

```

```

PA      (REGG ) UNIV CALIFORNIA.
PI      Billachone VM, Dillidine SL, Jolly DJ, Respass JG;
PI      Sandmeyer SB;
PI      WPI: 97-021229/02.
PT      New chimeric integrase for targeted vector integration in
PT      eukaryotic genomes - useful for gene therapy, providing more
PT      consistent gene expression and lower rates of insertional
PT      mutagenesis
PS      Claim 7: Page -: 98pp; English.
CC      The present sequence is a specific example of a chimeric integrase
CC      made up of three domains from MOMLV and Ty3, with at least one domain
CC      derived from Ty3 integrase. This protein can direct integration of a
CC      vector construct into a defined region of a target eukaryotic genome.
CC      As part of gene delivery vehicles and transduction competent
CC      recombinant retroviral particles it can be useful in somatic
CC      and germ cell gene therapy (in vivo or ex vivo) of a wide range of
CC      genetic, infectious, degenerative, inflammatory, cardiovascular and
CC      autoimmune diseases or cancer. Typical examples of the many potential
CC      applications include treatment of HIV infection, haemophilia, Alzheimer's
CC      disease and rheumatoid arthritis. This protein rather than wild-type
CC      retroviral integrase reduces the rate of insertional mutagenesis and the
CC      degree of variation in gene expression, and particularly does not disrupt
CC      tRNA gene expression.
CC      N.B. The present sequence is not shown in the specification, but is
CC      made up from the two original sequences as specified, MOMLV and Ty3,
CC      which are shown.
SQ      Sequence 429 AA:

Query Match      49.5%; Score 54; DB 1; Length 429;
Best Local Similarity 46.2%; Pred. No. 1.10e+02;
Matches      6; Conservative      3; Mismatches      4; Indels      0; Gaps      0;

DB      361 TKNLEPRMKGPGYT 373
      : : : : :
OY      2 ARTLNRYTGPYT 14

RESULT 11
ID      R0575 standard; protein: 431 AA.
AC      R0575;
DT      08-APR-1996 (first entry)
DE      Rat CRF2-beta receptor.
KW      CRF2-beta receptor; corticotropin-releasing factor-2 receptor;
KW      cerebrovascular disorder; memory disorder; Alzheimer disease.
OS      Rattus sp.
FH      key
      location/Qualifiers
FT      domain
FT      1. .117
FT      /label- Extracellular_N-terminal_domain
FT      118. .138
FT      /label- Transmembrane_domain
FT      139. .147
FT      /label- Intracellular_domain
FT      148. .167
FT      /label- Transmembrane_domain
FT      168. .184
FT      /label- Extracellular_domain
FT      185. .208
FT      /label- Transmembrane_domain
FT      209. .223
FT      /label- Intracellular_domain
FT      224. .244
FT      /label- Transmembrane_domain
FT      245. .261
FT      /label- Extracellular_domain
FT      262. .286
FT      /label- Transmembrane_domain
FT      287. .309
FT      /label- Intracellular_domain
FT      310. .329
FT      /label- Transmembrane_domain
FT      330. .342
FT      /label- Extracellular_domain
FT      343. .363

```


RESULT 14
 ID R75189 standard; Protein; 1196 AA.
 AC R75189;
 DT 30-MAY-1996 (first entry)
 DE Osteoinductive retrovirus RFB-14 pol gene product.
 KW RFB retrovirus; gag; pol; env; osteogenesis; osteoinductive protein;
 KW bone development; osteoporosis; gene therapy; polymerase;
 KW reverse transcriptase.
 OS Retrovirus RFB-14.
 PN DE4411718-A1.
 PD 12-OCT-1995.
 PF 05-APR-1994: 411718.
 PR 05-APR-1994: DE-411718
 PA (GSFU-) GSF FORSCHUNGSZENTRUM UNWELT & GESUNDEHEIT.
 PI Erile V, Gimbel W, Oestergaard M, Pedersen FS, Pedersen L;
 PI Schmidt J, Straus P;
 DR WPI: 95-352078/46.
 DR N-PSDB: 094266.
 PT RFB-14 retrovirus genome - and prodn. of osteo-inductive proteins
 PS Claim 14: Fig 1, 46pp; German.
 CC The full-length proviral genomic sequence of retrovirus RFB-14 has
 CC been determined. The virus codes for an osteoinductive protein,
 CC although the precise location of the coding region has not yet been
 CC identified. The virus may be useful in gene therapy of bone growth
 CC disorders such as osteoporosis. The present sequence is that of the
 CC viral pol gene product.
 SQ Sequence 1196 AA;

Query Match 49.5%; Score 54; DB 1; Length 1196;
 Best Local Similarity 46.2%; Pred. No. 1.10e+02;
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 1131 TKNEPRWKGPYT 1143
 : : : : :
 OY 2 ARTLNRYTGPYT 14

RESULT 15
 ID W00832 standard; Protein; 105 AA.
 DT 29-MAY-1997 (first entry)
 DE Variable light chain of anti-human Fas ligand antibody NOK-5.
 KW Variable region; light chain; human; Fas ligand; monoclonal;
 KW antibody; NOK-5; hybridoma; inhibition; apoptosis; assay;
 KW diagnosis; disease; hepatitis; infectious mononucleosis;
 KW systemic lupus erythematosus.
 OS Mus musculus.
 PN W09629350-A1.
 PD 26-SEP-1996.
 PF 21-MAR-1996: J00734.
 PR 20-MAR-1995: JP-087420.
 PR 27-OCT-1995: JP-303492.
 PA (SUME) SUMITOMO ELECTRIC IND CO.
 PI Kayagaki N, Nakata M, Okumura K, Yagita H;
 DR WPI: 96-443140/44.
 DR N-PSDB: T39558.
 PT Monoclonal antibody specifically recognising the Fas ligand - useful
 PT for the detection of Fas ligands either on cell surface or in
 PT solution
 PS Claim 28: Page 91; 133pp; Japanese.
 CC The present sequence is the light chain variable region of the
 CC anti-human Fas ligand monoclonal antibody (MAB) NOK-5. NOK-5 is
 CC produced by the hybridoma NOK-5 (FERM BP-5048), which was prepared
 CC by immunising mice with transformed human Fas ligand expressing COS
 CC cells, and fusing spleen cells isolated from the mice with myeloma
 CC P3x63Ag8.653 (ATCC CRL-1580) cells. The MAB recognises the human
 CC Fas ligand on the cell surface or in solution, and can be used to
 CC inhibit the apoptosis inducing cell surface Fas ligand/Fas
 CC reaction. The MAB can also be used for a Fas ligand assay in
 CC biological samples (e.g. human blood), especially for disease
 CC diagnosis, e.g. hepatitis, infectious mononucleosis and systemic
 CC lupus erythematosus.

SQ Sequence 105 AA;

Query Match 48.6%; Score 53; DB 1; Length 105;
 Best Local Similarity 62.5%; Pred. No. 1.40e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 91 HYSSPYTF 98
 : : : : :
 OY 8 RYTGPTTF 15

Search completed: Sat May 13 07:29:39 2000
 Job time : 7 secs.

 W E R E H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 07:22:27 2000: MspPar time 96.51 Seconds
 4.733 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-991-628-6
 Description: (1-15) from US08991628.pep
 Perfect Score: 109
 Sequence: 1 SARTLNRRYTGPTTF 15

Scoring table:
 Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database: swiss-prot38
 1:swissprot

Statistics: Mean 27.244; Variance 33.691; scale 0.809

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|----------------|-----------|
| 1 | 109 | 100.0 | 999 | 1 | DSG3_HUMAN | 8.31e-13 |
| 2 | 65 | 59.6 | 431 | 1 | IM44_YEAST | 5.59e-02 |
| 3 | 62 | 56.9 | 391 | 1 | GABG_XENLA | 2.41e-01 |
| 4 | 60 | 55.0 | 273 | 1 | HM01_CAEEL | 6.23e-01 |
| 5 | 60 | 55.0 | 334 | 1 | VG08_BPT4 | 6.23e-01 |
| 6 | 59 | 54.1 | 304 | 1 | CBP2_SIMVI | 9.94e-01 |
| 7 | 58 | 53.2 | 444 | 1 | NIFN_ANASP | 1.58e+00 |
| 8 | 58 | 53.2 | 469 | 1 | NIFD_MEITH | 1.58e+00 |
| 9 | 58 | 53.2 | 1237 | 1 | P0L4_DROME | 1.58e+00 |
| 10 | 57 | 52.3 | 387 | 1 | GAT6_CHICK | 2.45e+00 |
| 11 | 57 | 52.3 | 1013 | 1 | CHSA_EMENT | 2.45e+00 |
| 12 | 56 | 51.4 | 360 | 1 | IDH1_YEAST | 3.90e+00 |
| 13 | 56 | 51.4 | 1317 | 1 | NUC1OTRAT | 3.90e+00 |
| 14 | 55 | 50.5 | 245 | 1 | Y44C_RHISN | 6.08e+00 |
| 15 | 55 | 50.5 | 261 | 1 | GVPEFL_PROTEIN | 6.08e+00 |
| 16 | 55 | 50.5 | 391 | 1 | GABA_XENLA | 6.08e+00 |
| 17 | 55 | 50.5 | 431 | 1 | CRF2_MOUSE | 6.08e+00 |
| 18 | 55 | 50.5 | 433 | 1 | CBPA_ANOGE | 6.08e+00 |
| 19 | 55 | 50.5 | 455 | 1 | KR2_EBV | 6.08e+00 |
| 20 | 54 | 49.5 | 120 | 1 | P0L1_MOUSE | 9.41e+00 |
| 21 | 54 | 49.5 | 130 | 1 | V29K_TRYTC | 9.41e+00 |
| 22 | 54 | 49.5 | 194 | 1 | VMA3_VACCV | 9.41e+00 |
| 23 | 54 | 49.5 | 231 | 1 | RECO_COXBU | 9.41e+00 |

| | | | | | | | |
|----|----|------|------|---|------------|------------------------|----------|
| 24 | 54 | 49.5 | 252 | 1 | V29K_TRYSY | 29 KD PROTEIN | 9.41e+00 |
| 25 | 54 | 49.5 | 282 | 1 | POL_MLYCB | POL POLYPROTEIN [CONTA | 9.41e+00 |
| 26 | 54 | 49.5 | 365 | 1 | GAL7_YEAST | GALACTOSE-1-PHOSPHATE | 9.41e+00 |
| 27 | 54 | 49.5 | 367 | 1 | V669_METJA | PUTATIVE ATP-DEPENDENT | 9.41e+00 |
| 28 | 54 | 49.5 | 370 | 1 | GPRA_RAT | PROBABLE G PROTEIN-COU | 9.41e+00 |
| 29 | 54 | 49.5 | 390 | 1 | P0L3_MOUSE | RETROVIRUS-RELATED POL | 9.41e+00 |
| 30 | 54 | 49.5 | 428 | 1 | YUPL_YEAST | PROBABLE MANNOXYLTRANS | 9.41e+00 |
| 31 | 54 | 49.5 | 581 | 1 | POL_MLYRK | POL POLYPROTEIN [CONTA | 9.41e+00 |
| 32 | 54 | 49.5 | 746 | 1 | YLS8_CAEEL | HYPOTHETICAL 83.6 KD P | 9.41e+00 |
| 33 | 54 | 49.5 | 843 | 1 | POL_MLYAK | POL POLYPROTEIN [CONTA | 9.41e+00 |
| 34 | 54 | 49.5 | 1196 | 1 | POL_MLYVD | POL POLYPROTEIN [CONTA | 9.41e+00 |
| 35 | 54 | 49.5 | 1196 | 1 | POL_MLYAV | POL POLYPROTEIN [CONTA | 9.41e+00 |
| 36 | 54 | 49.5 | 1199 | 1 | POL_MLYVO | POL POLYPROTEIN [CONTA | 9.41e+00 |
| 37 | 54 | 49.5 | 1204 | 1 | POL_MLYVF | POL POLYPROTEIN [CONTA | 9.41e+00 |
| 38 | 54 | 49.5 | 1204 | 1 | POL_MLYV5 | POL POLYPROTEIN [CONTA | 9.41e+00 |
| 39 | 54 | 49.5 | 1204 | 1 | POL_MLYVF | POL POLYPROTEIN [CONTA | 9.41e+00 |
| 40 | 53 | 48.6 | 109 | 1 | Y85A_METJA | HYPOTHETICAL PROTEIN M | 1.45e+01 |
| 41 | 53 | 48.6 | 228 | 1 | Y85C_EMENT | HYPOTHETICAL 25.4 KD P | 1.45e+01 |
| 42 | 53 | 48.6 | 245 | 1 | YFIO_ECOLI | HYPOTHETICAL 27.8 KD L | 1.45e+01 |
| 43 | 53 | 48.6 | 2205 | 1 | P0LN_RUBVT | NONSTRUCTURAL POLYPROT | 1.45e+01 |
| 44 | 53 | 48.6 | 2261 | 1 | RRPL_MUMPM | RNA POLYMERASE BETA SU | 1.45e+01 |
| 45 | 53 | 48.6 | 2769 | 1 | THYG_BOVIN | THYOGLOBULIN PRECURSO | 1.45e+01 |

ALIGNMENTS

RESULT 1
 ID DSG3_HUMAN STANDARD: PRT: 999 AA.

AC P32926;
 DT 01-OCT-1993 (Rel. 27, created)
 DT 01-OCT-1993 (Rel. 27, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP MEDLINE: 92069753.
 RA Amagai M., Klaus-Kovtun V., Stanley J.R.;
 RT "Autoantibodies against a novel epithelial cadherin in pemphigus
 vulgaris, a disease of cell adhesion.";
 RL Cell 67:869-877(1991).

- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.
 INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
 FILAMENTS MEDIATING CELL-CELL ADHESION.
 CC -1 SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1 TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND
 CARCINOMAS.

- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 (POTENTIAL).
 CC -1 DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN
 DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE
 LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES
 AGAINST DSG3.

- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: M76482; AAA60230.1; -
 CC PIR: A41088; ITHUG3.
 CC HSSP: P09803; IEDH.
 CC MIM: 169655; -
 CC PROSITE: PS00232; CADHERIN_3.
 CC PFMW: PF00028; cadherin; 4.
 CC Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;

KW Calcium-Binding: Repeat.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 49 POTENTIAL.
 FT CHAIN 50 999 DESMOGLEIN 3.
 FT DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).
 FT TRANSEM 616 640 POTENTIAL.
 FT DOMAIN 641 999 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 50 158 CADHERIN 1.
 FT REPEAT 159 268 CADHERIN 2.
 FT REPEAT 269 383 CADHERIN 3.
 FT REPEAT 386 499 CADHERIN 4.
 FT REPEAT 910 935 DESMOGLEIN REPEAT 1.
 FT REPEAT 936 966 DESMOGLEIN REPEAT 2.
 FT CARBOHYD 110 110 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 459 459 POTENTIAL.
 FT CARBOHYD 545 545 POTENTIAL.
 SQ SEQUENCE 999 AA: 107503 MW: 60479DD46AC219A1 CRC64;
 Query Match 100.0% Score 109; DB 1; Length 999;
 Best Local Similarity 100.0% Pred. No. 8, 31e-13;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 512 SARTLNRYTGPYTF 526
 1 SARTLNRYTGPYTF 15

RESULT 2
 ID IM44_YEAST STANDARD: PRT: 431 AA.
 AC 001852;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE MITOCHONDRIAL IMPORT INNER MEMBRANE TRANSLOCASE SUBUNIT TIM44
 DE PRECURSOR (MITOCHONDRIAL PROTEIN 1) (INNER MEMBRANE
 DE IMPORT SITE PROTEIN 45) (ISPA45) (MEMBRANE IMPORT MACHINERY PROTEIN
 DE TIM44).
 GN TIM44 OR MP11 OR MIM44 OR ISPA45 OR YIL022M.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AB320;
 RX MEDLINE: 93010953.
 RA Mease A.C., Blom J., Grivell L.A., Melzer M.;
 RT "Mpi1, an essential gene encoding a mitochondrial membrane protein,
 RT is possibly involved in protein import into yeast mitochondria."
 RL EMO J. 11:3619-3628(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copley T., Dear S., Devlin K., Fraser A.,
 RA Guelcher S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Welsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RX MEDLINE: 93345448.
 RA Horst M., Jenoe P., Kronidou N.G., Bolliger L., Oppliger W.,
 RA Scherer P., Manning-Krieg U., Jascut T., Schatz G.;
 RT "Protein import into yeast mitochondria: the inner membrane import
 RT site protein ISPA45 is the Mpi1 gene product."
 RL EMO J. 12:3035-3041(1993).
 CC -1- FUNCTION: INVOLVED IN PROTEIN IMPORT INTO THE MITOCHONDRION.
 CC -1- PROBABLY INVOLVED IN TRANSLLOCATION ACROSS THE INNER MEMBRANE.
 CC AS A BINDING PROTEIN REQUIRED FOR DRIVING THE IMPORT OF
 CC PREPROTEINS. RECRUITS MITOCHONDRIAL HSP70 AND ITS CO-CHAPERONE
 CC (MGE1) TO DRIVE PROTEIN TRANSLLOCATION INTO THE MATRIX USING ATP

CC AS AN ENERGY SOURCE.
 CC -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
 CC LEAST 3 DIFFERENT PROTEINS (TIM17, TIM23, TIM44).
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE TIM44 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 246881, CA86970.1; -;
 CC EMBL: X67276; CA447693.1; -;
 CC PIR: S25196; S25196.
 CC SCD: L0001138; MPI1.
 CC Mitochondrion; Inner membrane; Transport; Protein transport;
 KW Translocation; Transit peptide; ATP-binding.
 FT TRANSIT 1 431 MITOCHONDRION.
 FT CHAIN ? 431 MITOCHONDRIAL IMPORT INNER MEMBRANE
 FT NP_BIND ? 431 ATP (POTENTIAL).
 SQ SEQUENCE 431 AA: 48854 MW: 8E98C738178B1E44 CRC64;
 Query Match 59.6% Score 65; DB 1; Length 431;
 Best Local Similarity 57.1% Pred. No. 5, 39e-02;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 13 SSRTLTARYSOYT 26
 1 SARTLNRYTGPYTF 14

RESULT 3
 ID GAB6_XENLA STANDARD: PRT: 391 AA.
 AC P70005;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TRANSCRIPTION FACTOR XGATA-6B (GATA BINDING FACTOR-6B).
 DE GATA-6A OR GATA-6.
 GN Xenopus laevis (African clawed frog).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LIVER;
 RX MEDLINE: 97180937.
 RA Gove C.D., Malmley M., Nijjar S., Bertwistle D., Gullie M.,
 RA Partridge G., Bomford A., Patient R.;
 RT "Over-expression of GATA-6 in Xenopus embryos blocks differentiation
 RT of heart precursors."
 RL EMO J. 16:355-368(1997).
 CC -1- FUNCTION: ASSOCIATED WITH CARDIAC SPECIFICATION AND CAN REGULATE
 CC CARDIAC-SPECIFIC TRANSCRIPTION DURING EMBRYOGENESIS. ACTIVATES
 CC THE EXPRESSION OF CARDIAC MYC-ALPHA IN VIVO.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y08865; CA470088.1; -;
 CC HSSP: P16768; 1GAU
 CC PRINTS: PR00619; GATA2NFINGER.

DR PROSITE: PS00344: GATA_ZN_FINGER: 2.
DR PFAM: PF00320: GATA: 2.
KW Transcription regulation: Activator: DNA-binding; Zinc-finger:
KM Nuclear protein.
FT ZN_FING 182 206 GATA-TYPE.
FT ZN_FING 236 260 GATA-TYPE.
FT DOMAIN 70 79 POLY-SER.
FT DOMAIN 241 245 POLY-THR.
FT DOMAIN 293 298 POLY-SER.
SQ SEQUENCE 391 AA: 41503 MW: C93C0CDE246204D3 CRC64:
Query Match 56.9%: Score 62; DB 1; Length 391;
Best Local Similarity 61.5%: Pred. No. 2,41e-01;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Db 120 ARPLNRYGSPY 132
QY 2 ARTLNRYGPT 14
RESULT 4
ID HMD1_CABEL STANDARD: PRT: 273 AA.
AC Q18273:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PUTATIVE HOMEBOX PROTEIN C28A5.4.
GN C28A5.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabdilitida;
OC Rhabdilitida; Rhabdilitidae; Rhabdilitinae; Peloderiinae; Caenorhabditis.
RN [1]
RC STRAIN-BRISTOL N2;
RA Palmer S.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEBOX
CC PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: Z32680; CAA83601.1; -;
DR HSSP: P02836; 1HDD.
DR WORMPEP: C28A5.4; CE05325.
DR PRINTS: PR00024; HOMEBOX.
DR PRINTS: PR00031; HTHREPRESSR.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
DR PFAM: PF00046; homeobox; 1.
KW Hypothetical protein: Homeobox: DNA-binding; Nuclear protein.
FT DNA_BIND 102 161 HOMEBOX.
SQ SEQUENCE 273 AA: 30163 MW: 1B93JF08C2576EB CRC64:
Query Match 55.0%: Score 60; DB 1; Length 273;
Best Local Similarity 40.0%: Pred. No. 6,23e-01;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
Db 215 PMTLEQYASPYLX 229
QY 1 SARTLNRYGPT 15
RESULT 5
ID VGO8_BPT4 STANDARD: PRT: 334 AA.
AC P19062:
DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE BASEPLATE STRUCTURAL PROTEIN GP8.
GN 8.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
OC T4-like phages.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D;
RX MEDLINE: 90384864.
RA Efimov V.P., Prilipov A.G., Mesyanzhinov V.V.;
RT "Nucleotide sequences of bacteriophage T4 genes 6, 7 and 8."
RL Nucleic Acids Res. 18:5313-5315(1990).
CC -1- FUNCTION: STRUCTURAL COMPONENT OF THE BASEPLATE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X15907; CAA34023.1; -;
DR PIR: J00658; G8BPT4.
DR PIR: JF0058; JF0058.
KW Structural protein.
SQ SEQUENCE 334 AA: 38008 MW: 4997860773E14899 CRC64:
Query Match 55.0%: Score 60; DB 1; Length 334;
Best Local Similarity 60.0%: Pred. No. 6,23e-01;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 97 DTRYPDPYTF 106
QY 6 NNRYGPT 15
RESULT 6
ID CBP2_SIMV1 STANDARD: PRT: 304 AA.
AC P42788:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE ZINC CARBOXYPEPTIDASE (EC 3.4.17.-) (FRAGMENT).
OS Simulium vittatum (Black fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Simuliidae; Simulium.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-GUT;
RX MEDLINE: 94093864.
RA Ramos A., Mahowald A., Jacobs-Lorena M.;
RT "Gut-specific genes from the black fly Simulium vittatum encoding
RT trypsin-like and carboxypeptidase-like proteins."
RL Insect Mol. Biol. 1:149-163(1993).
CC -1- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.
CC -1- TISSUE SPECIFICITY: GUT-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: L08481; AAA18531.1; -;
DR HSSP: P00730; 4CPA.

```

DR PROSITE: PS00132: CARBOXYPEPT_2N.1; 1.
DR PROSITE: PS00133: CARBOXYPEPT_2N.2; 1.
DR PFAM: PF00246: Zn_carboxypept; 1.
KW Hydrolase: Carboxypeptidase; Zinc.
FT METAL 58 58 ZINC (BY SIMILARITY).
FT METAL 61 61 ZINC (BY SIMILARITY).
FT METAL 184 184 ZINC (BY SIMILARITY).
FT ACT_SITE 236 236 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 259 259 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 125 148 BY SIMILARITY.
SQ SEQUENCE 304 AA: 34849 MW: 26633FBA6A9144 CRC64:
Query Match 54.1% Score 59; DB 1; Length 304;
Best Local Similarity 50.0%; Pred. No. 9.94e-01;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 214 AARSLAORVGTDT 227
OY 1 SARTLNRRYTGPT 14

RESULT 7
ID NIFN_AMASP STANDARD; PRT; 444 AA.
AC 044145;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE NITROGENASE IRON-MOLYBDENUM COFACTOR BIOSYNTHESIS PROTEIN NIFN.
GN NIFN.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria: Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
RN [1]
RP SEQUENCE FROM N.A.
RA Birkema W.J., Scarpino L.A., Haselkorn R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN THE BIOSYNTHESIS OF
CC THE PROSTHETIC GROUP OF NITROGENASE (FEMO COFACTOR).
CC -!- PATHWAY: FEMO COFACTOR BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U47055; AAA87948.1; -
DR PROSITE: PS00699; NITROGENASE_1.1; 1.
DR PROSITE: PS00090; NITROGENASE_1.2; FALSE_NEG.
DR PFAM: PF00148; oxidored_nitro; 1.
KW Nitrogen fixation.
SQ SEQUENCE 444 AA: 48332 MW: 766848616959037 CRC64:
Query Match 53.2% Score 58; DB 1; Length 444;
Best Local Similarity 46.2%; Pred. No. 1.58e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 245 AARLEDRFNIPY 257
OY 1 SARTLNRRYTGPT 13

RESULT 8
ID NIFD_MERTH STANDARD; PRT; 469 AA.
AC 027605;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NITROGENASE MOLYBDENUM-IRON PROTEIN ALPHA CHAIN (EC 1.18.6.1)
DE (NITROGENASE COMPONENT 1) (DINITROGENASE).

```

```

GN NIFD OR MTH1563.
OS Methanobacterium thermoautotrophicum.
OC Archaea: Euryarchaeota: Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H.
RX MEDLINE: 98037514.
RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-W., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakey D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pochler B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patel D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-T., Rice P., Nollig J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT delta: functional analysis and comparative genomics."
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- FUNCTION: THE KEY ENZYMIC REACTIONS IN NITROGEN FIXATION ARE
CC CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE
CC IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.
CC -!- CATALYTIC ACTIVITY: 8 REDUCED FERREDOXIN + 8 H(+) + N(2) + 16 ATP
CC -!- 8 OXIDIZED FERREDOXIN + 2 NH(3) + 16 ADP + 16 ORTHOPHOSPHATE.
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains that binds
CC 30-32 Fe, 2 Mo, and inorganic sulfur.
CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000916; AAB86037.1; -
DR HSSP: P00467; IMIO.
DR PROSITE: PS00699; NITROGENASE_1.1; FALSE_NEG.
DR PROSITE: PS00090; NITROGENASE_1.2; 1.
DR PFAM: PF00148; oxidored_nitro; 1.
KW Oxidoreductase; Nitrogen fixation; Molybdenum; Iron-sulfur.
SQ SEQUENCE 469 AA: 52866 MW: 7D2CD5FED9EC7B8 CRC64:
Query Match 53.2% Score 58; DB 1; Length 469;
Best Local Similarity 58.3%; Pred. No. 1.58e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 273 ARLLEERYNVNPY 284
OY 2 ARTLNRRYTGPT 13

RESULT 9
ID POL4_DROME STANDARD; PRT; 1237 AA.
AC P10394;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RETROVIRUS-RELATED POLYPROTEIN FROM TRANSPOSON 412 [CONTAINS:
DE PROTEASE (EC 3.4.23.-); REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE ENDONUCLEASE].
GN POL.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86274717.
RA Yuki S., Inouye S., Ishimaru S., Saigo K.;
RT "Nucleotide sequence characterization of a Drosophila
RT retictransposon, 412."
RL Eur. J. Biochem. 158:403-410(1986).

```


SO SEQUENCE 1317 AA: 145660 MW: 59399086B553030 CRC64:

Query Match 51.4% Score 56; DB 1: Length 1317;

Best Local Similarity 46.2% Pred. No. 3.90e+00;

Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 1145 NAOALMDRYGNY 1157

QY 1 SARTLNRYTGPY 13

RESULT 14

ID GVPL ANAFL STANDARD; PRT; 245 AA.

AC P55149;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUL-1998 (Rel. 35, Last annotation update)

DE GVPE/L PROTEIN.

GN GVPE/L.

OS Anabaena flos-aquae.

OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CCAP 1403/13F;

RX MEDLINE: 97217685.

RA Kinsman R., Hayes P.K.;

RT Genes encoding proteins homologous to halobacterial Gyps N, J, K, F

RT 6 L are located downstream of gypc in the cyanobacterium Anabaena

RT flos-aquae.

RT DNA Seq. 7:97-106(1997).

CC -I- FUNCTION: MAY PLAY A STRUCTURAL OR REGULATORY ROLE IN GAS

CC VESICLE SYNTHESIS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U17109; AAA58714.1; -

KW Gas vesicle.

SO SEQUENCE 245 AA: 28676 MW: 5CF0BFC999E11BC0 CRC64;

Query Match 50.5% Score 55; DB 1: Length 245;

Best Local Similarity 53.8% Pred. No. 6.08e+00;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 228 RIRYNNFTAPYTF 240

QY 3 RTLNNRYTGPYTF 15

RESULT 15

ID Y4ZC-RHISN STANDARD; PRT; 261 AA.

AC P55730;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE HYPOTHEICAL 28.3 KD PROTEIN Y4ZC.

GN Y4ZC.

OS Rhizobium sp. (strain NGR234).

OC Plasmid sym pNGR234a.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

CC Rhizobiaceae; Rhizobium.

CC -I- SIMILARITY: TO P.SYRINGAE (PV. PHASEOLICOLA) AVIRULENCE PROTEIN

CC AVIRPPH3.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AE000109; AAB91961.1; -

KW Hypothetical protein; Plasmid.

SO SEQUENCE 261 AA: 28349 MW: 52939C1919DFEC43 CRC64;

Query Match 50.5% Score 55; DB 1: Length 261;

Best Local Similarity 60.0% Pred. No. 6.08e+00;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 238 RSLANRYSNP 247

QY 3 RTLNNRYTGP 12

Search completed: Sat May 13 07:24:15 2000

Job time : 108 secs.

| Accession | Score | Length | Indels | Gaps |
|--|-------|--------|--------|------|
| CC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha; | | | | |
| OC Tetraodontiformes; Tetraodontidae; Fugu. | | | | |
| RN [1] | | | | |
| RP SEQUENCE FROM N.A. | | | | |
| RK MEDLINE: 98362517. | | | | |
| RA FOLTER R.T.M., BUTLER M.I.; | | | | |
| RT "A retrotransposon family from the pufferfish (fugu) <i>Fugu rubripes</i> ."; | | | | |
| RL Gene 215:241-249(1998). | | | | |
| DR EMBL: AF030881; AAC33526.1; - | | | | |
| DR PPM: PF00078; fvt; 1. | | | | |
| KM Polyprotein. | | | | |
| FT NON_TER | | | | |
| SO SEQUENCE 1187 AA; 132431 MW; 7AE7F01 CRC32; | | | | |
| Db 1055 SSRKLNPRDGPYT 1069 | | | | |
| Oy 1 SARLNNRYTGPYTF 15 | | | | |
| RESULT 3 | | | | |
| ID 024780. PRELIMINARY; PRT; 980 AA. | | | | |
| AC 024780. | | | | |
| DT 01-JAN-1998 (TREMblrel. 05. Created) | | | | |
| DT 01-JAN-1998 (TREMblrel. 05. Last sequence update) | | | | |
| DE 01-AUG-1998 (TREMblrel. 07. Last annotation update) | | | | |
| CD CELLULOXYRIN PHOSPHORYLASE. | | | | |
| GN CDP. | | | | |
| OS Clostridium thermocellum. | | | | |
| OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; | | | | |
| NC Clostridium. | | | | |
| RN [1] | | | | |
| RP SEQUENCE FROM N.A. | | | | |
| RC STRAIN-ATCC27405. | | | | |
| RA KANAGUCHI T., IREUCHI Y., TSUTSUMI N., KAN A., SUMITANI J., ARAI M.; | | | | |
| RL J. Ferment. Bioeng. 85:144-149(1998). | | | | |
| DR EMBL: AB006822; BAA22081.1; - | | | | |
| SO SEQUENCE 980 AA; 111183 MW; ABBAE30C CRC32; | | | | |
| Db 713 ARVLINRYKDSYTY 727 | | | | |
| Oy 2 ARTLNNRYT-GPYTF 15 | | | | |
| RESULT 4 | | | | |
| ID 09XE44. PRELIMINARY; PRT; 949 AA. | | | | |
| AC 09XE44. | | | | |
| DT 01-NOV-1999 (TREMblrel. 12. Created) | | | | |
| DT 01-NOV-1999 (TREMblrel. 12. Last sequence update) | | | | |
| DT 01-NOV-1999 (TREMblrel. 12. Last annotation update) | | | | |
| DE PUTATIVE REVERSE TRANSCRIPTASE. | | | | |
| GN TM22. | | | | |
| OS Arabidopsis thaliana (Mouse-ear cress). | | | | |
| OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; | | | | |
| OC eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons; | | | | |
| OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; | | | | |
| OC Arabidopsis. | | | | |
| RN [1] | | | | |
| RP SEQUENCE FROM N.A. | | | | |
| RC STRAIN-CV. COLUMBIA. | | | | |
| RA LIN X., KAUL S., SHEA T.P., FUJII C.Y., SHEN M., VANAKEN S.E., | | | | |
| BA BANGSTAD M.E., MASON T.M., BOWMAN C.L., RONNING C.M., BENITO M., | | | | |
| RA CARRERA A.J., CHEAST T.H., BOELL C.R., TOWN C.D., NIERMAN W.C., | | | | |
| RA FRASER C.M., VENTER J.C.; | | | | |
| RT "Arabidopsis thaliana chromosome II BAC TM2 genomic sequence."; | | | | |
| RL submitted (JUN-1999) to the EMBL/GenBank/DBJ databases. | | | | |

```

DR EMBL: AC007730; AAD37020.1; -.
SQ RNA-directed DNA polymerase.
Query Match 55.0%; Score 60; DB 10; Length 949;
Best Local Similarity 53.8%; Pred. No. 2.29e+00;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
SQ SEQUENCE 949 AA; 108395 MW; 6F3A521A CRC32;

Db 826 SRKRLSPRYGVP 868
| : | : | | | |
OY 1 SARLNNRYTGPY 13

RESULT 5 PRELIMINARY; PRT; 973 AA.
ID 081508
AC 081508:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE T7M24.4 PROTEIN.
T7M24.4.
GN T7M24.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eukaryliophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-CV. COLUMBIA.
RC HARMON G., LANGSTON Y., STONEKING T., DRONE K., AMES M.;
RA "The sequence of Arabidopsis thaliana T7M24."
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN-CV. COLUMBIA.
RC WASHU:
RA "The Arabidopsis thaliana Genome Sequencing Project."
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN-CV. COLUMBIA.
RC WATERSTON R.;
RA Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF077408; AAC26251.1; -.
DR PFAM: PF00078; tvc. 1.
SQ SEQUENCE 973 AA; 112369 MW; F6086EE2 CRC32;

Query Match 55.0%; Score 60; DB 10; Length 973;
Best Local Similarity 53.8%; Pred. No. 2.29e+00;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Db 826 SRKRLSPRYGVP 838
| : | : | | | |
OY 1 SARLNNRYTGPY 13

RESULT 6 PRELIMINARY; PRT; 163 AA.
ID 053240
AC 053240:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE HYPOTHEICAL 18.2 KD PROTEIN.
MY012.05.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
[1]
RN SEQUENCE FROM N.A.
RP OLIVER K., HARRIS D.;
RA Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]

```

RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RA BARRELL B.G., RAJANDREAM M.A., PARKHILL J., COLE S.T.;
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE: 96181548.
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPIELLA L.,
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
 RA COLE S.T.;
 RT "An integrated map of the genome of the tubercle bacillus,
 RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
 RT lepreae";
 RT Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
 DR EMBL: AL021287; CA16076.1;
 KW Hypothetical protein.
 SO SEQUENCE 163 AA; 18204 MW; D33CE4BB CRC32;

Query Match 54.1%; Score 59; DB 2; Length 163;
 Best Local Similarity 87.5%; Pred. No. 3.53e+00;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 111 ERYTGPYT 118
 OY 7 NRYTGPYT 14

RESULT 7
 ID 059397 PRELIMINARY; PRT; 330 AA.
 AC 059397;
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
 DE 330AA LONG HYPOTHEICAL N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE.
 GN PH1720.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OT3;
 RX MEDLINE: 98344137.
 RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
 RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOGAWA A., NAGAI Y.,
 RA SAKAI M., OGURA K., OTUWA R., NAKAZAWA H., TAKAMITA M., OHFUKU Y.,
 RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KOSHIDA N., OGUCHI A.,
 RA AOI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
 RA KIKUCHI H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
 RL DNA Res. 5:55-76(1998).
 DR EMBL: AP000007; BAA30834.1;
 DR PFM: PFO118; Semialdehyde; 1.
 SO SEQUENCE 330 AA; 37168 MW; D80618ED CRC32;

Query Match 54.1%; Score 59; DB 1; Length 330;
 Best Local Similarity 53.8%; Pred. No. 3.53e+00;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 50 GGLNRTSDYNF 62
 OY 3 RLNNRYTGPYTF 15

RESULT 8
 ID 050378 PRELIMINARY; PRT; 360 AA.
 AC 050378;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
 DE FERRIC EXOCHELIN BIOSYNTHESIS.
 GN FXA.
 OS Mycobacterium smegmatis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MC6;
 RX MEDLINE: 95191405.
 RA FISS E.H.;
 RT "Identification of genes involved in the sequestration of iron in
 RT mycobacteria: the ferric exochelin biosynthetic and uptake pathways.";
 RT Mol. Microbiol. 14:557-569(1994).
 DR EMBL: U10425; AAC43261.1;
 DR HSSP: P23882; 1FMT.
 DR PFM: PFO0551; formyl-transf. 1.
 SO SEQUENCE 360 AA; 41130 MW; A81D7F3B CRC32;

Query Match 54.1%; Score 59; DB 2; Length 360;
 Best Local Similarity 38.5%; Pred. No. 3.53e+00;
 Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

DB 241 RALSDPYPNATY 253
 OY 3 RLNNRYTGPYTF 15

RESULT 9
 ID 066695 PRELIMINARY; PRT; 531 AA.
 AC 066695;
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE HYPOTHEICAL 62.8 KD PROTEIN.
 GN AQ.372.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RX MEDLINE: 98196666.
 RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus";
 RL Nature 392:353-358(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AE000686; AAC06652.1;
 KW Hypothetical protein.
 SO SEQUENCE 531 AA; 62830 MW; 585EBE7E CRC32;

Query Match 54.1%; Score 59; DB 2; Length 531;
 Best Local Similarity 66.7%; Pred. No. 3.53e+00;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 80 TLKNRLNPYTF 91
 OY 4 TLNNRYTGPYTF 15

RESULT 10
 ID 001124 PRELIMINARY; PRT; 211 AA.
 AC 001124;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE COAT PROTEIN VP1 (FRAGMENT).
 GN ID.
 OS foot and mouth disease virus C.

OC Viruses: ssRNA positive-strand viruses, no DNA stage: Picornaviridae;
 OC Aphthovirus.
 [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 92260626.
 RA MARTINEZ M.A., DOPAZO J., HERNANDEZ J., MATEU M.G., SOBRINO F.,
 RA DOMINGO E., KNOWLES N.J.;
 RT "Evolution of the capsid protein genes of foot-and-mouth disease
 RT virus: antigenic variation without accumulation of amino acid
 RT substitutions over six decades.";
 RL J. Virol. 66:357-365(1992).
 CC -1- SUBMIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- MISCELLANEOUS: THE COAT PROTEIN VP1 CONTAINS THE MAIN ANTIGENIC
 CC DETERMINANTS OF THE VIRION; THEREFORE, CHANGES IN ITS SEQUENCE
 CC MUST BE RESPONSIBLE FOR THE HIGH ANTIGENIC VARIABILITY OF THE
 CC VIRUS.
 DR EMBL: M90377: AAA91495.1: -.
 DR PFAM: PF00073: rhv: 1.
 KM COAT protein.
 FT NON_TER 1 1
 FT DOMAIN 134 157 ANTIGENIC A (BY SIMILARITY).
 FT DOMAIN 188 211 ANTIGENIC C (BY SIMILARITY).
 FT NON_TER 211 211
 SQ SEQUENCE 211 AA: 23028 MW: 81474E6 CRC32:
 Query Match 53.2%; Score 58; DB 14; Length 211;
 Best Local Similarity 53.8%; Pred. No. 5.43e+00;
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 Db 125 RVLTARYTGTTY 137
 Oy 3 RLNNRYTGPTY 15
 RESULT 11
 ID 029749 PRELIMINARY: PRT: 332 AA.
 AC 029749;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-AUG-1998 (TEMBLrel. 07, Last annotation update)
 DE NITRATE REDUCTASE, GAMMA SUBUNIT, PUTATIVE.
 GN AF0501.
 OS Archaeoglobus fulgidus.
 OC Archaea: Euryarchaeota: Archaeoglobales: Archaeoglobaceae;
 OC Archaeoglobus.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE: 98049343.
 RA KLEIN H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C.,
 RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON C.G., GILL S.,
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOTUS B.,
 RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLOER A., ZHOU L.,
 RA OVERBEER K., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
 RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
 RA SADDON P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
 RA VENTER J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 DR EMBL: AE001069: AAB90736.1: -.
 DR TIGR: AF0501: -.
 KW Hypothetical protein.
 SO SEQUENCE 332 AA: 38386 MW: 6844D788 CRC32:
 Query Match 53.2%; Score 58; DB 1; Length 332;
 Best Local Similarity 50.0%; Pred. No. 5.43e+00;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 53 RTTYRFDSPYT 64
 Oy 3 RLNNRYTGPTY 14
 RESULT 12
 ID 017381 PRELIMINARY: PRT: 506 AA.
 AC 017381;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
 DE CE-FKH-1.
 GN CE-FKH-1.
 OS Caenorhabditis elegans.
 OC Eukaryota: Metazoa: Nematoda: Secernentea: Rhabditia: Rhabditidae;
 OC Rhabditina: Rhabditidae; Rhabditidae; Pelodetrinae; Caenorhabditis.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-N2;
 RA AZZARIA M., GOSZCZYNSKI B., CHUNG M.A., KALB J.M., MCGHEE J.D.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MORTIMORE B.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U51163: AAA96319.1: -.
 DR EMBL: 292833: CAB07378.1: -.
 DR HSP: Q63245; ZFH.
 DR PROSITE: PS00657; FORK_HEAD_1: 1.
 DR PROSITE: PS00658; FORK_HEAD_2: 1.
 DR PFAM: PF00250; Fork_head: 1.
 DR PRINTS: PR00053; FORKHEAD.
 SO SEQUENCE 506 AA: 55426 MW: 54119AA3 CRC32:
 Query Match 53.2%; Score 58; DB 5; Length 506;
 Best Local Similarity 50.0%; Pred. No. 5.43e+00;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 443 TLNQVAPLYL 454
 Oy 4 TLNNRYTGPTY 15
 RESULT 13
 ID 007117 PRELIMINARY: PRT: 770 AA.
 AC 007117;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
 DE HYPOTHEICAL 87.2 KD PROTEIN.
 OS Barley stripe mosaic virus (BSMV).
 OC Viruses: ssRNA positive-strand viruses, no DNA stage: Hordeivirinae.
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 9006400.
 RA KOZLOV YU.V., AFANAS'EV B.N., RUPASOV V.V., GOLOVA YU.B.,
 RA KULIAVA O.I., DOLYA V.V., ATABEKOV I.G., BAEV A.A.;
 RT "Primary structure of RNA 3 of barley stripe mosaic virus and its
 RT variability.";
 RL Mol. Biol. (Mosk) 23:1080-1090(1989).
 DR EMBL: X52774: CAA36983.1: -.
 DR PFAM: PF00978: RNA_dep_RNApol2: 1.
 KW Hypothetical protein.
 SO SEQUENCE 770 AA: 87176 MW: 34477E66 CRC32:
 Query Match 53.2%; Score 58; DB 14; Length 770;
 Best Local Similarity 50.0%; Pred. No. 5.43e+00;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 720 KALNNRYSPYS 731
 Oy 3 RLNNRYTGPTY 14

```

RESULT 14
ID 076025 PRELIMINARY: PRT: 80 AA.
AC 076025;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DE 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94107601.
RA MURPHY E., KORBER B.T., GEORGES-COURBOT M.C., YU B., PINTER A.,
RA COOK D., KIENY M.P., GEORGES A., MATHIOT C., BARRE-SINOSSI F.,
RA GIRARD M.;
RT "Diversity of V3 region sequences of human immunodeficiency viruses
RT type 1 from the central African Republic.";
RL AIDS Res. Hum. Retroviruses 9:997-1006(1993).
DR EMBL: L11496; AAC37829.1; -.
DR PFAM: PF00516; GP120; 1.
KM Envelope protein.
FT NON_TER
SO SEQUENCE 80 AA; 9239 MM; 15E98296 CRC32;

Query Match 52.3% Score 57: DB 14: Length 80:
Best Local Similarity 70.0% Pred. No. 8.32e+00;
Matches 7: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

Db 63 NTRYTGSYTL 72
Oy 6 NNRITGPTF 15

RESULT 15
ID 005776 PRELIMINARY: PRT: 175 AA.
AC 005776;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DE 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE SIMILAR TO MSF1 PROTEIN.
GN L8167.12.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C (AB972);
RX MEDLINE: 97313267.
RA JOHNSTON M., HILLER L., RILES L., ALBERMANN K., ANDRE B., ANSORGE W.,
RA BENES V., BRUCKNER M., DELIUS H., DUBOIS E., DUSTERHOFT A.,
RA ENTIAN K.D., FLOETH M., GOFEAU A., HEBLING U., HEUMANN K.,
RA HEUSS-NEITZEL D., HILBERT H., HILGER F., KLEINE K., KOTTER P.,
RA LOUIS E.J., MESSENGUY F., MEWES H.W., MIOGA T., MOSTL D.,
RA MULLER-AUER S., NENTWICH U., OBERMAIER B., PIRAVANDI E., POHL T.M.,
RA PORTFELLE D., PURNELLE B., RECHMANN S., RIEGER M., RINKE M., ROSE M.,
RA SCHARFE M., SCHERENS B., SCHOLLER P., SCHWAGER C., SCHWARZ S.,
RA UNDERWOOD A.P., URRESTARAZU L.A., VANDENBOL M., VERHASSELT P.,
RA VIERENDELS F., VOET M., VOLCKAERT G., VOSS H., WAMBUIT R., WEDLER E.,
RA WEDLER H., ZIMMERMANN F.K., ZOLLNER A., HANI J., HOEISEL J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C (AB972);
RA PAULEY A.;
RN Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C (AB972);
RA WATERSTON R.;

```

```

RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C (AB972);
RA CHERRY J.M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U14913; AAB67434.1; -.
SO SEQUENCE 175 AA; 20108 MM; 9FCA0042 CRC32;

Query Match 52.3% Score 57: DB 3: Length 175:
Best Local Similarity 46.2% Pred. No. 8.32e+00;
Matches 6: Conservative 4: Mismatches 3: Indels 0: Gaps 0:

Db 19 SRAFNRYPNPYS 31
Oy 2 ARTLNRYTGPYT 14

Search completed: Sat May 13 07:28:45 2000
Job time : 251 secs.

```

THIS PAGE BLANK (USPTO)

 M E R E N E
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 07:37:01 2000; Maspar time 3.01 Seconds
 118.060 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-991-628-7
 Description: (1-15) from US08991628.pep
 Perfect Score: 99
 Sequence: 1 OSGTMRTRHSTGCTN 15

Scoring table:
 PAM 150
 Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-geneseq35
 1:geneseqp

Statistics: Mean 17.183; Variance 54.742; scale 0.314

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | ID | Description | Pred. No. |
|------------|-------|--------------------|------|--------------------------------------|-----------|
| 1 | 99 | 100.0 | 15 | W04847 Self epitope of desmog | 5.70e-04 |
| 2 | 99 | 100.0 | 999 | R30742 Human pemphigus vulgar | 5.70e-04 |
| 3 | 54 | 54.5 | 572 | R06023 Viral haemagglutinin n | 6.17e+01 |
| 4 | 49 | 49.5 | 140 | R10423 Noe L gene product. | 1.99e+02 |
| 5 | 49 | 49.5 | 392 | R03182 Topoisomerase II bindi | 1.99e+02 |
| 6 | 48 | 48.5 | 802 | R56550 Cold acclimatization p | 2.51e+02 |
| 7 | 48 | 48.5 | 830 | R80314 CITTA (class II transa | 2.51e+02 |
| 8 | 48 | 48.5 | 9872 | R57057 Class II trans activat | 2.51e+02 |
| 9 | 48 | 48.5 | 1085 | R61048 A nidulans phosphoenol | 2.51e+02 |
| 10 | 48 | 48.5 | 1106 | R80312 CITTA (class II transa | 2.51e+02 |
| 11 | 48 | 48.5 | 1130 | R72452 Class II transactivato | 2.51e+02 |
| 12 | 48 | 48.5 | 1130 | R81569 Class II transactivato | 2.51e+02 |
| 13 | 48 | 48.5 | 1130 | R57056 Class II trans activat | 2.51e+02 |
| 14 | 48 | 48.5 | 1130 | R80313 CITTA (class II transa | 2.51e+02 |
| 15 | 48 | 48.5 | 1207 | R80311 CITTA (class II transa | 2.51e+02 |
| 16 | 47 | 47.5 | 483 | R80214 Bacillus licheniformis | 3.16e+02 |
| 17 | 47 | 47.5 | 483 | R88449 Alpha-amylase H1331/N1 | 3.16e+02 |
| 18 | 47 | 47.5 | 483 | R25434 Alpha-amylase mutant H | 3.16e+02 |
| 19 | 47 | 47.5 | 483 | R10579 R10579 Mutant alpha-amylase s | 3.16e+02 |
| 20 | 47 | 47.5 | 483 | R88450 Alpha-amylase H1331/T1 | 3.16e+02 |
| 21 | 47 | 47.5 | 483 | R88451 Alpha-amylase V1286/H1 | 3.16e+02 |
| 22 | 47 | 47.5 | 483 | R58734 M15V alpha-amylase. | 3.16e+02 |
| 23 | 47 | 47.5 | 483 | W05232 Bacillus licheniformis | 3.16e+02 |

| | | | | | | |
|----|----|------|-----|---|--------------------------------|----------|
| 24 | 47 | 47.5 | 483 | 1 | R58011 M197P alpha-amylase. | 3.16e+02 |
| 25 | 47 | 47.5 | 483 | 1 | R58729 M15H alpha-amylase. | 3.16e+02 |
| 26 | 47 | 47.5 | 483 | 1 | R57986 M197T alpha-amylase. | 3.16e+02 |
| 27 | 47 | 47.5 | 483 | 1 | R58737 M15E alpha-amylase. | 3.16e+02 |
| 28 | 47 | 47.5 | 483 | 1 | M57982 Alpha-amylase mutant M | 3.16e+02 |
| 29 | 47 | 47.5 | 483 | 1 | M57981 Alpha-amylase mutant M | 3.16e+02 |
| 30 | 47 | 47.5 | 483 | 1 | M73510 Alpha-amylase protein | 3.16e+02 |
| 31 | 47 | 47.5 | 483 | 1 | M58728 M15D alpha-amylase. | 3.16e+02 |
| 32 | 47 | 47.5 | 483 | 1 | R58730 M15K alpha-amylase. | 3.16e+02 |
| 33 | 47 | 47.5 | 483 | 1 | R58727 M15N alpha-amylase. | 3.16e+02 |
| 34 | 47 | 47.5 | 483 | 1 | R58725 M15A alpha-amylase. | 3.16e+02 |
| 35 | 47 | 47.5 | 483 | 1 | R58008 M197F alpha-amylase. | 3.16e+02 |
| 36 | 47 | 47.5 | 483 | 1 | R88442 Alpha-amylase M104D/H1 | 3.16e+02 |
| 37 | 47 | 47.5 | 483 | 1 | M08204 B. licheniformis mutant | 3.16e+02 |
| 38 | 47 | 47.5 | 487 | 1 | M80215 M4 form alpha-amylase | 3.16e+02 |
| 39 | 47 | 47.5 | 487 | 1 | W05238 Bacillus licheniformis | 3.16e+02 |
| 40 | 47 | 47.5 | 511 | 1 | Y01582 An alpha-amylase (Am-L | 3.16e+02 |
| 41 | 47 | 47.5 | 511 | 1 | Y01582 An alpha-amylase prote | 3.16e+02 |
| 42 | 47 | 47.5 | 512 | 1 | Y00769 B. licheniformis alpha | 3.16e+02 |
| 43 | 47 | 47.5 | 512 | 1 | M73509 Alpha-amylase protein. | 3.16e+02 |
| 44 | 47 | 47.5 | 512 | 1 | M10325 Alpha-amylase acid | 3.16e+02 |
| 45 | 47 | 47.5 | 531 | 1 | R98007 Pelb signal-fermamy1-1 | 3.16e+02 |

ALIGNMENTS

RESULT 1
 ID W04847; standard; peptide: 15 AA.
 AC W04847;
 DT 18-FEB-1997 (first entry)
 DE Self epitope of desmoglein 3, implicated in autoimmune disease.
 KW Tolerisation; self-epitope; antigen; autoimmune disease.
 KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
 KW pemphigus vulgaris; desmoglein; multiple sclerosis;
 KW herpes simplex virus; adenovirus; phosphomannomutase;
 KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
 KW influenza; haemagglutinin; reovirus; sigma protein.
 OS Homo sapiens.
 PN W09627387-A1.
 PD 12-SEP-1996.
 PE 07-MAR-1996; U03182.
 PR 07-MAR-1995; US-400796.
 PA (HARD) HARVARD COLLEGE.
 PI Strominger JL, Wucherpfennig KW;
 DR WPI: 96-425218/42.
 PT pemphigus vulgaris auto-antigens and multiple sclerosis non-self
 PT antigens - useful in disease treatment, and method for
 PT identification of other self and non-self antigens implicated in
 PT auto-immune disease
 PS Claim 1: Page 42: 58pp: English.
 CC Pharmaceutical preparations for tolerisation to antigens comprise
 CC either an isolated human non-collagen or non-mysin basic protein
 CC (MBP) polypeptide which is capable of tolerising an individual to an
 CC autoantigen; or an isolated human pathogen polypeptide capable of
 CC tolerising an individual to that polypeptide. In both cases, the
 CC polypeptide (whether self or non-self) includes an amino acid
 CC sequence corresponding to a sequence motif for a MHC class II
 CC protein, such as HLA-DR, which is associated with a human autoimmune
 CC disease and which binds to the polypeptide to activate autoreactive
 CC T-cells in individuals with the autoimmune disease. This peptide is
 CC derived from the human desmoglein 3 protein (amino acids 762-786)
 CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
 CC derived from the human desmoglein protein are described in W04847-47.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 99; DB 1; Length 15;
 Best local Similarity 100.0%; Pred. No. 5.70e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 OSGTMRTRHSTGCTN 15
 |||||||||||||
 Qy 1 OSGTMRTRHSTGCTN 15

| RESULT | 2 |
|--------|--|
| ID | R30742 standard: Protein; 999 AA. |
| AC | R30742.1 |
| DT | 14-JUN-1993 (first entry) |
| DE | Human pemphigus vulgaris 130kD antigen. |
| KW | Pemphigus vulgaris; skin disease; autoantibodies; |
| KW | keratinocyte cell surface antigen; glycoprotein; cell |
| OS | Homo sapiens. |
| PN | US798918-A. |
| PD | 15-DEC-1992. |
| PF | 27-NOV-1991; 798918. |
| PR | 27-NOV-1991; US-798918. |
| PA | (USSH) US DEPT HEALTH & HUMAN SERVICE. |
| PI | Amagai M, Kleihs-Kovtun V, Stanley JR; |
| DR | WPI: 93-0673436/08. |
| DR | N-PSDB: Q25992. |
| FT | DNA encoding pemphigus vulgaris antigen - useful in proteins for |
| FT | diagnostic and therapeutic uses |
| PS | Disclosure; Fig 7; 50pp: English. |
| CC | This sequence is the pemphigus vulgaris 130kD antigen. The protein |
| CC | and its encoding DNA may be used in the diagnosis and treatment of |
| CC | pemphigus vulgaris. It is thought that the antigen may be a cell |
| CC | adhesion molecule. |
| CC | Sequence 999 AA; |
| 50 | |

```

Query Match      100.0%;   Score 99;   DB 1;   Length 999;
Best Local Similarity 100.0%;   Pred. No. 5.70e-04;
Matches      15;   Conservative      0;   Mismatches      0;   Indels      0;   Gaps      0;

```

| Db | 762 | QSGTMRTRHSTGCTN | 776 |
|----|-----|-----------------|-----|
| | | | |
| Qy | 1 | QSGTMRTRHSTGCTN | 15 |

| RESULT | 3 | standard: protein: 572 AA. |
|--------|--|----------------------------|
| ID | R06023 | |
| AC | R06023: | |
| DT | 04-DEC-1990 | (first entry) |
| DE | Viral haemagglutinin neuraminidase. | |
| KW | Haemagglutinin neuraminidase; vaccine; parainfluenza: ds. | |
| OS | Bovine parainfluenza type III virus. | |
| PN | J02156883-A. | |
| PD | 15-JUN-1990. | |
| PF | 09-DEC-1988: 311655. | |
| PR | 09-DEC-1988: JP-311655. | |
| PA | (JAPG) NIPPON ZEON KK. | |
| DR | WPI: 90-228484/30. | |
| DR | N-PSDB: Q05319. | |
| PT | Recombinant vaccinia virus - is which all or part of DNA coding | |
| PT | membrane fusion protein in combined to genom region. | |
| PS | disclosure: P: Japanese. | |
| CC | Fragment of parainfluenza viral membrane fusion protein encoding DNA | |
| CC | may be incorporated into the vaccina virus, which may then be used | |
| CC | as a live vaccine 572 AA: | |
| Q | Sequence 572 AA: | |

| | | | | |
|-----------------------|--------|---------------------|-------|---------------------------------|
| Query Match | 54.58; | Score 54; | DB 1; | Length 572; |
| Best Local Similarity | 53.38; | Pred. No. 6.17e+01; | | |
| Matches | 8; | Conservative | 4; | Mismatches 3; Indels 0; Gaps 0; |

```
Db      16 EPGTTDRDRHSSKATN 30
      ::||| | |||: |||
Oy      1 QSGTMRTRHSTGCTN 15
```

RESULT 4
ID R10423 standard; Protein; 140 AA.
AC R10423:
DT 09-APR-1991 (first entry)
DE Mod L gene product.
KW Mod genes; parasite; toxin; promoter; modulation; flavones; mod L
KW *Bradyrhizobium japonicum*.

PN US4983519-A.
PD 08-JAN-1991.
PF 24-JUL-1987; 077561.
PR 24-JUL-1987; US-077561.
PA (SRAC/) STACEY G.
PI Stacey G, Schnell MG, Nieuwkoop AJ, Deshmame NA, Banfalvi Z:
DR WPI; 91-036225/05.
DR N-PSDB; Q10325.
PT Nod genes and control sequences of *Bradyrhizobium japonicum* -
PT used for improving nodulation and for prodn. of foreign protein,
PT e.g. plant parasite toxin
PR
PS Disclosure; Page 4; 11pp; English.
CC The nod I gene product may be used to restore nodulating ability
CC to Nod- mutants of *Bradyrhizobium* or *Rhizobium* strains and can
CC enhance nodulating ability.
SQ Sequence 140 AA;

| | | | | |
|-----------------------|--------|---------------------|-------|---------------|
| Query Match | 49.58; | Score 49; | DB 1; | Length 140; |
| Best Local Similarity | 50.08; | Pred. No. 1.99e+02; | | |
| Matches | 6; | Conservative | 5; | Mismatches 1; |
| | | | | Indels 0; |
| | | | | Gaps 0; |

| | | | |
|----|----|---------------|----|
| Db | 85 | SKRRRHATAGTK | 96 |
| | | ::: :: :: : | |
| QY | 4 | TMRTIRHSTGCTN | 15 |

RESULT 5
 ID Y03182 standard: protein; 392 AA.
 AC Y03182;
 DT 16-JUN-1999 (first entry)
 DE Topoisomerase II binding protein 1.
 KW Topoisomerase II binding protein; TopBP; anticancer agent.
 OS Homo sapiens.
 PN J11075856-A.
 PD 23-MAR-1999.
 PF 17-SEP-1997; 251544.
 PR 17-SEP-1997; JP-251544.
 PA (CHUS) CHUGAI PHARM CO LTD.
 PA (TSUR/) TSURUO T.
 DR WPI: 99-257704/22.
 DR N-PSDB: X28153.
 PT New Topoisomerase II - binding protein - useful as an anticancer agent
 PT agent
 PS Claim 2; Page 13-14; 28pp; Japanese.
 CC This sequence represents the topoisomerase II binding protein (TopBP) of the invention. The TopBP protein is useful as an anticancer agent. TopBPB can be used as the target molecule for anticancer agent.
 CC Sequence 392 AA;
 CQ

| | | | | |
|-----------------------|--------|---------------------|--------|---------------|
| Query Match | 49.5%; | Score 49; | DB 1; | Length 392; |
| Best Local Similarity | 66.7%; | Pred. No. 1.99e+02; | | |
| Matches | 6; | Conservative | 1; | Mismatches 2; |
| | | | Indels | 0; |
| | | | Gaps | 0; |

Db 1 MRRHDGTGA 9
||| |
QY 5 MRRHSTGG 13

RESULT 6
ID R56550 standard; Protein: 802 AA.
AC R56550;
DT 20-FEB-1995 (first entry)
DE Cold acclimation protein CAPI60 from spinach leaf.
KW Cold acclimation; cold-tolerance; transgenic plant; drought-
KM drought-resistance; drought-tolerance; crop improvement; CAPI60;
KW freezing-tolerance.
OS Spinacia oleracea.
PN W09417186-A.
PD 04-AUG-1994.
PF 21-JAN-1994; 000581.
PR 21-JAN-1993; US-007107.
PA (UYFL) UNIV FLORIDA.
PI Guy CL, Haskell DW, Hoffig A, Neven LG;

DR WPI: 94-264100/32.
 DR N-PSDB: Q70904.
 PT Cold acclimatization proteins CAP65 and CAP160 and their
 PT nucleotide sequences - used to confer increased cold tolerance
 PT and drought resistance on plants and microorganisms by genetic
 PT transformation
 PS Claim 6: Page 29; 49pp; English.
 CC Transgenic plants expressing this cold acclimatization
 CC protein have improved cold tolerance and/or drought-resistance.
 CC The cell expressing the protein is from the family Solanaceae,
 CC a citrus plant, a bacterium or a yeast cell.
 SO Sequence 802 AA.

Query Match 48.5%; Score 48; DB 1; Length 802;
 Best Local Similarity 53.3%; Pred. No. 2.51e+02;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

DB 768 OPGTRRHITOGTE 782
 1 OSGTMRHSTGCTN 15

RESULT 7
 ID W80314 standard; Protein: 830 AA.
 AC W80314;
 DT 01-FEB-1999 (first entry)
 DE ClrTA (class II transactivator) type IV protein.
 KM ClrTA gene; class II transactivator; MHC class II molecule;
 KM Interferon-gamma; Interleukin-4; vaccine; cancer treatment.
 OS Homo sapiens.
 PN EP-874049-A1.
 PD 28-OCT-1998.
 PE 21-APR-1998; 400968
 PR 22-APR-1997; FR-004954.
 PA (TRGE) TRANSGENE SA.
 PI Mach B;
 DI WPI: 98-559115/48.
 DR DNA encoding MHC class II trans-activator polypeptide(s) - useful
 PT e.g. as primers for enzymatic amplification, as detection probes or
 PT as inhibitors of expression of genes encoding MHC in dendritic cells
 PS Disclosure: Pages 66-69; 86pp; French.
 CC The present sequence represents type IV ClrTA (class II transactivator)
 CC protein. The products can be used to treat diseases for which
 CC enhanced expression of genes coding for MHC class II molecules is
 CC desired, especially where the enhanced expression is desired in
 CC dendritic cells or after induction by a cytokine, especially
 CC interferon-gamma or interleukin-4. Inhibitors of the gene or protein
 CC can be used to treat diseases for which reduced expression of genes
 CC coding for MHC class II molecules is desired or can be used as vaccines,
 CC especially for cancer treatment.
 SO Sequence 830 AA.

Query Match 48.5%; Score 48; DB 1; Length 830;
 Best Local Similarity 55.6%; Pred. No. 2.51e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 497 OPGTLRARQ 505
 1 OSGTMRH 9

RESULT 8
 ID W57057 standard; Protein: 987 AA.
 AC W57057;
 DT 17-AUG-1998 (first entry)
 DE Class II trans activator (ClrTA) 151 deletion mutant.
 KM Class II trans activator; MHC; ClrTA; autoimmune disease; treatment;
 KM transplantation; xenograft; major histocompatibility complex; diabetes;
 KM gene therapy; arthritis; ribozyme; 151 deletion mutant.
 OS Homo sapiens.
 OS Synthetic.
 FT Key Location/Qualifiers
 FT Misc_difference 6

FT /note- "wild type 143 amino acid residues deletion
 after this position".
 FT Misc_difference 7 /label- E1501
 FT /note- "wild type Glu is replaced by Ile"
 FT Misc_difference 8 /label- L1511
 FT /note- "wild type Leu is replaced by Ile"
 PN W09815626-A2.
 PD 16-APR-1998.
 PE 08-OCT-1997; G02751.
 PR 21-MAR-1997; GB-005911.
 PR 08-OCT-1995; GB-020940.
 PA (CHIL-) INST CHILD HEALTH.
 PI Fabre JW, Gustafsson KT, Yun S;
 DI WPI: 98-240813/21.
 DR N-PSDB: V28717.
 PT Deleted form of the class II trans-activator that reduces class II
 PT antigen expression - and ribozyme directed to trans-activator mRNA,
 PT related nucleic acid, vectors, transformed cells and antibodies,
 PT used for treating auto-immune disease and to inhibit xeno-graft
 PT rejection
 PS Claim 15: Pages 69-72; 104pp; English.
 CC This represents a 151 deletion mutant of the class II trans activator
 CC (ClrTA) polypeptide. The N-terminal region of the wild type ClrTA is
 CC deleted and can be used for reducing the expression of major
 CC histocompatibility complex (MHC) class II antigens. A ribozyme targeted
 CC to bases 1159-1161 (GUA) of human ClrTA mRNA, or corresponding target in
 CC other species and the ClrTA polypeptide can be used to reduce expression
 CC of MHC class II antigens, particularly for treatment of autoimmune
 CC disease (e.g. arthritis and diabetes) or to treat non-human animals
 CC intended as source of xenografts. The ribozyme and the ClrTA polypeptide
 CC may be generated in vivo by gene therapy, using the encoding nucleic
 CC acids targeted for localised suppression of the immune response. Material
 CC from transgenic animals in which at least some cells are stably
 CC transfected with the ClrTA encoding nucleic acids are used for animal-co-
 CC class II antigens. In cells that express them constitutively or after
 CC lymphokine induction. It has no transcription activity but still binds to
 CC regulatory proteins in the promoter region of class II genes, so
 CC competitively inhibits endogenous ClrTA.
 SO Sequence 987 AA.

Query Match 48.5%; Score 48; DB 1; Length 987;
 Best Local Similarity 55.6%; Pred. No. 2.51e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 654 OPGTLRARQ 662
 1 OSGTMRH 9

RESULT 9
 ID P61048 standard; Protein: 1085 AA.
 AC P61048;
 DT 03-SEP-1991 (first entry)
 DE A.nidulans phenol pyruvic acid carboxylase gene product.
 KM PPC; photosynthesis; pal81; pal171; pal172; pal172A.
 OS Anacystis nidulans.
 PN J61056082-A.
 PD 20-MAR-1986.
 PE 24-AUG-1984; 177352.
 PR 24-AUG-1984; JP-177352.
 PA (KOHZ/) KOHZUKI H.
 DI WPI: 86-115948/18.
 DR N-PSDB: N60838.
 PT PPC gene of cloned photosynthetic microorganisms - for
 PT intensifying photosynthesis carbon di-oxide-fixing ability of
 PT bacteria.
 PS Disclosure: Fig 1; 16pp; Japanese.
 CC Transforming a bacteria with the PPC encoding sequence intensifies
 CC CO2 fixation ability. Other PPC genes may also be found by
 CC hybridisation with an A.nidulans derived probe.

SO Sequence 1085 AA:
 Query Match 48.5%; Score 48; DB 1; Length 1085;
 Best Local Similarity 54.5%; Pred. No. 2.51e+02;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1054 TGLMRSRYSKG 1064
 : : : : :
 2 SGTMRHSTG 12

RESULT 10
 ID W80312 standard; Protein: 1106 AA.
 AC W80312;
 DT 01-FEB-1999 (first entry)
 DE C17A (class II transactivator) type II protein.
 KW C17A gene; class II transactivator; MHC class II molecule;
 KM Interferon-gamma; Interleukin-4; vaccine; cancer treatment.
 OS Homo sapiens.
 PN EP-874049-A1.
 PD 28-Oct-1998.
 PF 21-Apr-1998; 400968.
 PS 22-Apr-1997; FR-004954.
 PA (TRGE) TRANSGENE SA.
 PI Mach B:
 DR WPI: 98-55915/48.
 PT DNA encoding MHC class II trans-activator polypeptide(s) - useful
 e.g. as primers for enzymatic amplification, as detection probes or
 as inhibitors of expression of genes encoding MHC in dendritic cells
 PS Claim 10: Pages 58-61; 86pp; French.
 CC The present sequence represents type II C17A (class II transactivator)
 CC protein. The products can be used to treat diseases for which
 CC enhanced expression of genes coding for MHC class II molecules is
 CC desired, especially where the enhanced expression is desired in
 CC dendritic cells or after induction by a cytokine, especially
 CC interferon-gamma or interleukin-4. Inhibitors of the gene or protein
 CC can be used to treat diseases for which reduced expression of genes
 CC coding for MHC class II molecules is desired or can be used as vaccines,
 CC especially for cancer treatment.
 SO Sequence 1106 AA;

Query Match 48.5%; Score 48; DB 1; Length 1106;
 Best Local Similarity 55.6%; Pred. No. 2.51e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 773 QPGLRRARQ 781
 : : : : :
 1 QSGTMRTRH 9

RESULT 11
 ID R72452 standard; Protein: 1130 AA.
 AC R72452;
 DT 08-DEC-1995 (first entry)
 DE Class II transactivator protein.
 KW Class II transactivator protein; MHC class II;
 KM Insulin dependent diabetes; multiple sclerosis; lupus erythematosus;
 KW rheumatoid arthritis; Kozak box; ATP/GTP binding cassette;
 KM yeast RNA binding protein; transcription activation domain.
 OS Homo sapiens.
 PI Key
 DR WPI: 96-151325/15.
 PT Methods which inhibit transcription activation by C17A - causes
 PT inhibition of MHC class II gene expression, used in therapy of
 PT auto-immune disorders
 PS Example 1; Page 36-41; 64pp; English.
 CC A genomic DNA sequence (718028) codes for the class II
 CC transactivator, C17A (R81569), a protein essential for activation
 CC of transcription of MHC class II genes. The C17A transcription
 CC activation domain provides useful information for identifying cpds.
 CC which inhibit C17A-dependent transcription. Such cpds. are
 CC potential autoimmune disease therapeutics by virtue of their ability
 CC to inhibit transcription of the MHC class II genes. A second portion

PF 26-AUG-1994; 113378.
 PR 26-AUG-1993; EP-113665.
 PA (MACH/) MACH B F.
 PI Mach BF:
 DR WPI: 95-148715/20.
 DR N-PSDB: 088110.
 PT Gene encoding a protein displaying class II trans-activator
 PT activity - useful in compsn. for treating diseases associated
 PT with impaired expression of MHC-II genes, etc.
 PS Claim 11; Page 17-21; 32pp; English.
 CC This protein sequence is encoded by the gene represented in 088110.
 CC The protein is a class II transactivator (C17A) essential for the
 CC control of MHC class II gene expression in B lymphocytes. The DNA
 CC sequence has two in frame start codons. The first of these is at position
 CC 116 and acts as the translation initiation site. The second is present
 CC (at position 188) in the context of a perfect Kozak box and may also
 CC serve as a start codon for a protein of 1106 amino acids. The protein
 CC contains three stretches rich in proline/serine/threonine. It also
 CC contains a region rich in glutamate/aspartate (an acidic region) and an
 CC ATP/GTP binding cassette. The acidic regions followed by three stretches
 CC rich in proline, serine and threonine resembles a transcription
 CC activation domain. There is also a leucine-rich region around amino
 CC acids 979 to 1061 that shows weak homology with the N terminal portion of
 CC a yeast RNA binding protein. The protein can be used for the
 CC identification of inhibitors to proteins with C17A activity. The
 CC sequence can be used to treat diseases where a decrease in the level of
 CC expression of MHC cells II genes is desired e.g. Insulin dependent
 CC diabetes, multiple sclerosis, lupus erythematosus and Rheumatoid
 CC arthritis.
 SO Sequence 1130 AA;

Query Match 48.5%; Score 48; DB 1; Length 1130;
 Best Local Similarity 55.6%; Pred. No. 2.51e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 797 QPGLRRARQ 805
 : : : : :
 1 QSGTMRTRH 9

RESULT 12
 ID R81569 standard; Protein: 1130 AA.
 AC R81569;
 DT 22-MAY-1996 (first entry)
 DE Class II transactivator.
 KW C17A; class II transactivator; transcription; MHC;
 KM major histocompatibility complex class II; autoimmune disease;
 KM therapy.
 OS Homo sapiens.
 PI Key
 DR WPI: 96-151325/15.
 PT Methods which inhibit transcription activation by C17A - causes
 PT inhibition of MHC class II gene expression, used in therapy of
 PT auto-immune disorders
 PS Example 1; Page 36-41; 64pp; English.
 CC A genomic DNA sequence (718028) codes for the class II
 CC transactivator, C17A (R81569), a protein essential for activation
 CC of transcription of MHC class II genes. The C17A transcription
 CC activation domain provides useful information for identifying cpds.
 CC which inhibit C17A-dependent transcription. Such cpds. are
 CC potential autoimmune disease therapeutics by virtue of their ability
 CC to inhibit transcription of the MHC class II genes. A second portion

CC of CIITA, the interaction domain, mediates interaction of CIITA
CC and its target protein and activates MHC class II promoters.
CC Cpbds. that inhibit this binding are also potential autoimmune
CC disease therapeutics.
SO Sequence 1130 AA;

Query Match 48.5%; Score 48; DB 1; Length 1130;
Best Local Similarity 55.6%; Pred. No. 2.51e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
DB 797 QPGTLRARQ 805
1 QSGTMRTRH 9

RESULT 13
ID W57056 standard; Protein: 1130 AA.
AC W57056;
DT 17-AUG-1998 (first entry)
DE Class II trans activator (CIITA) polypeptide.
KW Class II trans activator; MHC; CIITA; autoimmune disease; treatment;
KW transplantation; xenograft; major histocompatibility complex; diabetes;
KW gene therapy; arthritis; ribozyme.
OS Homo sapiens.
FH Key
FT Region 52..137 Location/Qualifiers
FT Region /note="N-terminal region rich in glutamate/aspartate"
FT Region 163..195
FT Region /note="region rich in proline/serine/threonine"
FT Region 209..237
FT Region /note="region rich in proline/serine/threonine"
FT Region 261..322
FT Region /note="region rich in proline/serine/threonine"
PD W09815626-A2.
PN 16-APR-1998.
PF 08-OCT-1997; G02751.
PR 21-MAR-1997; GB-005911.
PR 08-OCT-1996; GB-020940.
PA (CHIL-) INST CHILD HEALTH.
PI Fabre JW, Gustafsson KT, Yun S;
DR WPI: 98-240813/21.
DR N-PSDB: V28716.
PT Deleted form of the class II trans-activator that reduces class II
PT antigen expression - and ribozyme directed to trans-activator mRNA,
PT related nucleic acid, vectors, transformed cells and antibodies,
PT used for treating auto-immune disease and to inhibit xeno-graft
PT rejection
PS Claim 14; Fig 5; 104pp; English.
CC This represents a class II trans activator (CIITA) polypeptide. When
CC the N-terminal region of the CIITA is deleted, it can be used for
CC reducing the expression of major histocompatibility complex (MHC) class
CC II antigens. A ribozyme targeted to bases 1159-1161 (GUA) of human CIITA
CC mRNA, or corresponding target in other species and the CIITA polypeptide
CC can be used to reduce expression of MHC class II antigens, particularly
CC for treatment of autoimmune disease (e.g. arthritis and diabetes) or to
CC treat non-human animals intended as source of xenografts. The ribozyme
CC and the CIITA polypeptide may be generated in vivo by gene therapy, using
CC the encoding nucleic acids targeted for localised suppression of the
CC immune response. Material from transgenic animals in which at least some
CC cells are stably transfected with the CIITA encoding nucleic acids are
CC used for animal-to-human transplantation. The CIITA polypeptide
CC suppresses synthesis of class II antigens, in cells that express them
CC constitutively or after lymphokine induction. It has no transcription
CC activity but still binds to regulatory proteins in the promoter region of
CC class II genes, so competitively inhibits endogenous CIITA.
SO Sequence 1130 AA;

Query Match 48.5%; Score 48; DB 1; Length 1130;
Best Local Similarity 55.6%; Pred. No. 2.51e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
DB 797 QPGTLRARQ 805
1 QSGTMRTRH 9

QY 1 QSGTMRTRH 9

RESULT 14
ID W80313 standard; Protein: 1130 AA.
AC W80313;
DT 01-FEB-1999 (first entry)
DE CIITA (class II transactivator) type III protein.
KW CIITA gene; class II transactivator; MHC class II molecule;
KW interferon-gamma; interleukin-4; vaccine; cancer treatment.
OS Homo sapiens.
FH Key
FT Region 52..137 Location/Qualifiers
FT Region /note="N-terminal region rich in glutamate/aspartate"
FT Region 163..195
FT Region /note="region rich in proline/serine/threonine"
FT Region 209..237
FT Region /note="region rich in proline/serine/threonine"
FT Region 261..322
FT Region /note="region rich in proline/serine/threonine"
PD W09815626-A2.
PN 16-APR-1998.
PF 08-OCT-1997; G02751.
PR 21-MAR-1997; GB-005911.
PR 08-OCT-1996; GB-020940.
PA (CHIL-) INST CHILD HEALTH.
PI Fabre JW, Gustafsson KT, Yun S;
DR WPI: 98-240813/21.
DR N-PSDB: V28716.
PT Deleted form of the class II trans-activator that reduces class II
PT antigen expression - and ribozyme directed to trans-activator mRNA,
PT related nucleic acid, vectors, transformed cells and antibodies,
PT used for treating auto-immune disease and to inhibit xeno-graft
PT rejection
PS Claim 14; Fig 5; 104pp; English.
CC This represents a class II trans activator (CIITA) polypeptide. When
CC the N-terminal region of the CIITA is deleted, it can be used for
CC reducing the expression of major histocompatibility complex (MHC) class
CC II antigens. A ribozyme targeted to bases 1159-1161 (GUA) of human CIITA
CC mRNA, or corresponding target in other species and the CIITA polypeptide
CC can be used to reduce expression of MHC class II antigens, particularly
CC for treatment of autoimmune disease (e.g. arthritis and diabetes) or to
CC treat non-human animals intended as source of xenografts. The ribozyme
CC and the CIITA polypeptide may be generated in vivo by gene therapy, using
CC the encoding nucleic acids targeted for localised suppression of the
CC immune response. Material from transgenic animals in which at least some
CC cells are stably transfected with the CIITA encoding nucleic acids are
CC used for animal-to-human transplantation. The CIITA polypeptide
CC suppresses synthesis of class II antigens, in cells that express them
CC constitutively or after lymphokine induction. It has no transcription
CC activity but still binds to regulatory proteins in the promoter region of
CC class II genes, so competitively inhibits endogenous CIITA.
SO Sequence 1130 AA;

Query Match 48.5%; Score 48; DB 1; Length 1130;
Best Local Similarity 55.6%; Pred. No. 2.51e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 797 QPGTLRARQ 805
1 QSGTMRTRH 9

RESULT 15
ID W80311 standard; Protein: 1207 AA.
AC W80311;
DT 01-FEB-1999 (first entry)
DE CIITA (class II transactivator) type I protein.
KW CIITA gene; class II transactivator; MHC class II molecule;
KW interferon-gamma; interleukin-4; vaccine; cancer treatment.
OS Homo sapiens.
FH Key
FT Region 52..137 Location/Qualifiers
FT Region /note="N-terminal region rich in glutamate/aspartate"
FT Region 163..195
FT Region /note="region rich in proline/serine/threonine"
FT Region 209..237
FT Region /note="region rich in proline/serine/threonine"
FT Region 261..322
FT Region /note="region rich in proline/serine/threonine"
PD W09815626-A2.
PN 16-APR-1998.
PF 08-OCT-1997; G02751.
PR 21-MAR-1997; GB-005911.
PR 08-OCT-1996; GB-020940.
PA (CHIL-) INST CHILD HEALTH.
PI Fabre JW, Gustafsson KT, Yun S;
DR WPI: 98-240813/21.
DR N-PSDB: V28716.
PT Deleted form of the class II trans-activator that reduces class II
PT antigen expression - and ribozyme directed to trans-activator mRNA,
PT related nucleic acid, vectors, transformed cells and antibodies,
PT used for treating auto-immune disease and to inhibit xeno-graft
PT rejection
PS Claim 14; Fig 5; 104pp; English.
CC This represents a class II trans activator (CIITA) polypeptide. When
CC the N-terminal region of the CIITA is deleted, it can be used for
CC reducing the expression of major histocompatibility complex (MHC) class
CC II antigens. A ribozyme targeted to bases 1159-1161 (GUA) of human CIITA
CC mRNA, or corresponding target in other species and the CIITA polypeptide
CC can be used to reduce expression of MHC class II antigens, particularly
CC for treatment of autoimmune disease (e.g. arthritis and diabetes) or to
CC treat non-human animals intended as source of xenografts. The ribozyme
CC and the CIITA polypeptide may be generated in vivo by gene therapy, using
CC the encoding nucleic acids targeted for localised suppression of the
CC immune response. Material from transgenic animals in which at least some
CC cells are stably transfected with the CIITA encoding nucleic acids are
CC used for animal-to-human transplantation. The CIITA polypeptide
CC suppresses synthesis of class II antigens, in cells that express them
CC constitutively or after lymphokine induction. It has no transcription
CC activity but still binds to regulatory proteins in the promoter region of
CC class II genes, so competitively inhibits endogenous CIITA.
SO Sequence 1207 AA;

Query Match 48.5%; Score 48; DB 1; Length 1207;
Best Local Similarity 55.6%; Pred. No. 2.51e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 55.6%; Pred. No. 2.51e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 874 QPGLRARQ 882

1:1:1:1:1:
QY 1 QSGTMRTRH 9

Search completed: Sat May 13 07:37:08 2000
Job time : 7 secs.

DB 762 OSGTMRHSTGCTN 776
|||||
OY 1 OSGTMRHSTGCTN 15

RESULT 2
ENTRY BVFSL #type complete
TITLE SOL protein, large splice form - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 22-Jun-1999

ACCESSIONS
REFERENCE A41146
#authors Delaney, S.J.; Hayward, D.C.; Barleben, F.; Fischbach, K.F.; Miklos, G.L.G.
#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:7214-7218
#title Molecular cloning and analysis of small optic lobes, a structural brain gene of Drosophila melanogaster.
#cross-references MUID:91334436
#accession A41146
#molecule_type mRNA

COMMENT #residues 1-1597 #label DEL
#cross-references GB:M64084; NID:q2760822; PIDN:AA895431.1; PID:q158483
The sol (small optic lobes) mutation eliminates certain classes of columnar neurons.
COMMENT An alternate splice form of 395 amino acids is observed, in which the first 393 are identical to the large sol protein.

GENETICS
#gene sol
#cross-references FlyBase:FBgn0003464
#map_position 19F4
CLASSIFICATION #superfamily sol protein; calpain catalytic domain homology
KEYWORDS #alternative splicing; brain; zinc finger
FEATURE
12-29 #region zinc finger CCCC motif
141-158 #region zinc finger CCCC motif
649-667 #region zinc finger CCCC motif
673-702 #region glutamine/histidine-rich
713-730 #region zinc finger CCCC motif
754-771 #region zinc finger CCCC motif
936-953 #region zinc finger CCCC motif
1047-1307 #domain calpain catalytic domain homology #label CALP
1082,1248,1268 #active_site Cys, His, Asn #status predicted
SUMMARY #length 1597 #molecular_weight 174713 #checksum 8253

Query Match 57.6%; Score 57; DB 1; Length 1597;
Best Local Similarity 61.5%; Pred. No. 3,67e-01;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 873 SGAIKPRHSTGGS 885
|||||
OY 2 SGTMRHSTGCT 14

RESULT 3
ENTRY T08428 #type complete
TITLE gene small optic lobes protein - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 20-Aug-1999

ACCESSIONS
REFERENCE T08428
#authors Malezka, R.; de Couet, H.G.; Miklos, G.L.
#journal Proc. Natl. Acad. Sci. U.S.A. (1998) 95:3731-3736
#title Data transferability from model organisms to human beings: insights from the functional genomics of the flightless region of Drosophila.
#cross-references MUID:98188272
#accession T08428
#status translated from GB/EMBL/DBJ

##molecule_type DNA
##residues 1-1597 #label MAL
#cross-references EMBL:AF01777; NID:q3004652; PID:q3004662
##experimental_source strain Cantons

GENETICS
#gene sol
#introns 81/2; 290/1; 594/2; 704/1; 1233/3; 1402/3; 1491/3; 1539/2
CLASSIFICATION #superfamily sol protein; calpain catalytic domain homology
FEATURE 1047-1307
SUMMARY #domain calpain catalytic domain homology #label CALP
#length 1597 #molecular_weight 174695 #checksum 8817

Query Match 57.6%; Score 57; DB 2; Length 1597;
Best Local Similarity 61.5%; Pred. No. 3,67e-01;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 873 SGAIKPRHSTGGS 885
|||||
OY 2 SGTMRHSTGCT 14

RESULT 4
ENTRY S33886 #type fragment
TITLE DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain
ORGANISM #formal_name Euplotes octocarinatus
DATE 08-Dec-1992 #sequence_revision 10-Nov-1995 #text_change 12-Sep-1997

ACCESSIONS
REFERENCE S33886; S70414; S73103
#authors Kaufmann, J.; Klein, A.
#journal Nucleic Acids Res. (1992) 20:4445-4450
#title Gene dosage as a possible major determinant for equal expression levels of genes encoding RNA polymerase subunits in the hypotrichous ciliate Euplotes octocarinatus.
#cross-references MUID:93027138
#accession S33886
#status translation not shown

##molecule_type DNA
##residues 1-478 #label KAUF
##residues 1-478 #label KAUF
#cross-references EMBL:X66452; NID:q9314; PID:q578408
REFERENCE S70412
#authors Kaufmann, J.; Florian, V.; Klein, A.
#journal Nucleic Acids Res. (1992) 20:5985-5989
#title TGA cysteine codons and intron sequences in conserved and nonconserved positions are found in macronuclear RNA polymerase genes of Euplotes octocarinatus.
#cross-references MUID:93096587
#accession S70414
##molecule_type DNA
##residues 1-124 #label KA2
#cross-references EMBL:X66452

GENETICS
#gene RPB1
#genetic_code GCG
#introns 30/3; 74/1
CLASSIFICATION #superfamily human DNA-directed RNA polymerase II largest chain
KEYWORDS DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc finger
SUMMARY #length 478 #checksum 3090

Query Match 56.6%; Score 56; DB 2; Length 478;
Best Local Similarity 53.8%; Pred. No. 6,02e-01;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 301 QAGQVNRHKSOG 313
|||||
OY 1 OSGTMRHSTGCT 13

RESULT 5
ENTRY A44991 #type complete

TITLE protein-tyrosine kinase (EC 2.7.1.112) fyn - mouse
 ALTERNATE_NAMES kinase-related transforming protein (fyn)
 ORGNISM #formal_name Mus musculus #common_name house mouse
 DATE 03-Jun-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999

ACCESSIONS A44991
 REFERENCE A44991
 #authors Cooke, M.P.; Perlmuter, R.M.
 #journal New Biol. (1989) 1:66-74
 #title Expression of a novel form of the fyn proto-oncogene in hematopoietic cells.
 #cross-references NID:91175680
 #accession A44991
 #status Preliminary
 #molecule_type mRNA
 #residues 1-534 #label COO
 #cross-references GB:M27266; NID:9193357; PIDN:AA37644.1; PID:9309241
 #note In the authors' translation an additional residue Leu was shown after Lys, for residue 459

CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
 KEYWORDS #homology: SH2 homology; SH3 homology
 #ATP: autophosphorylation; phosphoprotein; phosphotransferase; tyrosine-specific protein kinase

FEATURE
 89-138 #domain SH3 homology #label SH3\
 149-246 #domain protein kinase homology #label SH2\
 266-524 #domain protein kinase ATP-binding motif
 274-282 #region protein kinase ATP-binding motif
 SUMMARY #length 534 #molecular_weight 60057 #checksum 2574

Query Match
 Best Local Similarity 57.1%; Score 55; DB 2; Length 534;
 Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 72 HTGTRRGCTGVT 85
 :|||:|||||
 OY 1 QSGTMRTRHSTGCT 14

RESULT 6
 ENTRY S33568 #type complete
 TITLE protein-tyrosine kinase (EC 2.7.1.112) fyn - chicken
 ORGNISM #formal_name Gallus gallus #common_name chicken
 DATE 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 18-Jun-1999

ACCESSIONS S33568; S36351; S20807
 REFERENCE S33568
 #authors Sudol, M.; Graulich, H.; Newman, L.; Sarkar, A.; Sukeyawa, J.; Yamamoto, T.
 #journal Oncogene (1993) 8:823-831
 #title A novel yes-related kinase, Yrk, is expressed at elevated levels in neural and hematopoietic tissues.
 #cross-references NID:93205395
 #accession S33568
 #molecule_type mRNA
 #residues 1-534 #label SUDI
 #cross-references EMBL:X52841
 REFERENCE S36351
 #authors Sudol, M.
 #submission Submitted to the EMBL Data Library, May 1990
 #accession S36351
 #molecule_type mRNA
 #residues 1-348; 'I', 350-534 #label SUD2
 #cross-references EMBL:X52841; NID:962861; PIDN:CAA37025.1; PID:962862

GENERIC #gene fyn
 CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
 KEYWORDS #homology: SH2 homology; SH3 homology
 #ATP: phosphoprotein; phosphotransferase; proto-oncogene; transforming protein; tyrosine-specific protein kinase

FEATURE
 89-138 #domain SH3 homology #label SH3\
 149-246 #domain SH2 homology #label SH2\

266-524 #domain protein kinase homology #label KIN\
 274-282 #region protein kinase ATP-binding motif
 296 #binding_site ATP (Lys) #status predicted
 417-528 #binding_site phosphate (Tyr) (covalent) #status predicted

SUMMARY #length 534 #molecular_weight 60267 #checksum 4864

Query Match
 Best Local Similarity 57.1%; Score 55; DB 2; Length 534;
 Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 72 HTGTRRGCTGVT 85
 :|||:|||||
 OY 1 QSGTMRTRHSTGCT 14

RESULT 7
 ENTRY TVHUSY #type complete
 TITLE protein-tyrosine kinase (EC 2.7.1.112) fyn, splice form B - human
 ALTERNATE_NAMES #ynb; p59(fyn); protein-tyrosine kinase slk; protein-tyrosine kinase syn; proto-oncogene fyn
 ORGNISM #formal_name Homo sapiens #common_name man
 DATE 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 11-Jun-1999

ACCESSIONS A24314; A25389; I59120; I79512
 REFERENCE A24314
 #authors Semba, K.; Nishizawa, M.; Miyajima, N.; Yoshida, M.C.; Sukegawa, J.; Yamaneishi, Y.; Sasaki, M.; Yamamoto, T.; Toyoshima, K.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1986) 83:5459-5463
 #title yes-related protooncogene, syn, belongs to the protein-tyrosine kinase family.
 #cross-references NID:86287278
 #accession A24314
 #molecule_type mRNA
 #residues 1-537 #label SEM
 #cross-references GB:M4333; NID:9181171; PIDN:MAC08285.1; PID:9181172

REFERENCE A25389
 #authors Kawakami, T.; Pennington, C.Y.; Robbins, K.C.
 #journal Mol. Cell. Biol. (1986) 6:4195-4201
 #title Isolation and oncogenic potential of a novel human src-like gene.
 #cross-references NID:87089775
 #accession A25389
 #molecule_type mRNA
 #residues 1-183; 'S', 185-436; 'R', 438-537 #label KAM1
 #cross-references GB:M4676; NID:9338227; PIDN:AA36515.1; PID:9338228
 #note the authors translated the codon GAA for residue 265 as Gln and GGA for residue 278 as Gln

REFERENCE I59120
 #authors Kawakami, T.; Kawakami, Y.; Aaronson, S.A.; Robbins, K.C.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1988) 85:3870-3874
 #title Acquisition of transforming properties by fyn, a normal SRC-related human gene.
 #cross-references NID:88234523
 #accession I59120
 #status translated from GB/EMBL/DBJ
 #molecule_type DNA
 #residues 514-522 #label KAM2
 #cross-references GB:M20284; NID:9182842; PIDN:AA52491.1; PID:9182843
 #experimental_source clone clone PYN(C)-11; termination mutant p56(fyn)
 #accession I79512
 #status translated from GB/EMBL/DBJ
 #molecule_type DNA
 #residues 514-524; 'ORS' #label KAM3
 #cross-references GB:M20285; NID:9182844; PIDN:AA52492.1; PID:9182845
 #experimental_source clone clone PYN(C)-21; recombinant mutant p57(fyn)

REFERENCE A58826
 #authors Peters, D.J.; McGrew, B.R.; Perron, D.C.; Liptak, L.M.; Laudano, A.P.
 #journal Oncogene (1990) 5:1313-1319
 #title In vivo phosphorylation and membrane association of the fyn

proto-oncogene product in IM-9 human lymphoblasts.
#cross-references MUID:91016431
#contents annotation: myristylation: Tyr-531 phosphorylation
REFERENCE A51398
#authors Noble, M.; Musacchio, A.; Saraste, M.; Wierenga, R.
#submission submitted to the Brookhaven Protein Data Bank, May 1993
#cross-references PDB:1SHF
#contents annotation: X-ray crystallography, 1.9 angstroms, residues 84-142

REFERENCE A58827
#authors Noble, M.E.M.; Musacchio, A.; Saraste, M.; Courtneidge, S.A.; Wierenga, R.K.
#journal EMBO J. (1993) 12:2617-2624
#title Crystal structure of the SH3 domain in human Fyn: comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin.
#cross-references MUID:93327750
#contents annotation: X-ray crystallography, 1.9 angstroms
REFERENCE A65653
#authors Musacchio, A.; Saraste, M.; Wilmanns, M.
#submission submitted to the Brookhaven Protein Data Bank, May 1995
#cross-references PDB:1FYN
#contents annotation: X-ray crystallography, 2.3 angstroms, residues 81-142

REFERENCE A58828
#authors Musacchio, A.; Saraste, M.; Wilmanns, M.
#journal Nature Struct. Biol. (1994) 1:546-551
#title High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides.
#cross-references MUID:95393198
#contents annotation: X-ray crystallography, 2.3 angstroms
REFERENCE A66268
#authors Morton, C.J.; Pugh, D.J.R.; Campbell, I.D.
#submission submitted to the Brookhaven Protein Data Bank, April 1996
#cross-references PDB:1NYF
#contents annotation: conformation by (1)H-NMR, residues 84-141

REFERENCE A66269
#authors Morton, C.J.; Pugh, D.J.R.; Campbell, I.D.
#submission submitted to the Brookhaven Protein Data Bank, April 1996
#cross-references PDB:1NYG
#contents annotation: conformation by (1)H-NMR, residues 84-141

GENETICS
#gene GDB:FYN
#cross-references GDB:118797; OMIM:137025
#map_position 6q21-6q21
COMPLEX monomer
FUNCTION
#description catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
CLASSIFICATION
#superfamily protein-tyrosine kinase src: protein kinase
#homology: SH2 homology; SH3 homology
#formal_name: SH2 homology; SH3 homology
KEYWORDS
#alternative splicing: ATP; autophosphorylation; blocked amino end; lipoprotein; monomer; myristylation; phosphoprotein; phosphotransferase; proto-oncogene; thioester bond; transforming protein; tyrosine-specific protein kinase

FEATURE
89-138 #domain SH3 homology #label SH3\
149-246 #domain SH2 homology #label SH2\
269-527 #domain protein kinase homology #label KIN\
277-285 #region protein kinase ATP-binding motif\
#modified_site myristylated amino end (Gly) (1n mature form) #status predicted\
#binding_site palmitate (Cys) (covalent) #status predicted\
#binding_site palmitate (Cys) (covalent) #status predicted\
#active_site Lys #status predicted\
#binding_site phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted\
#binding_site phosphate (Tyr) (covalent) #status experimental

SUMMARY
#length 537 #molecular-weight 60762 #checksum 9747
Query Match 55.6%; Score 55; DB 1; Length 537;
Best Local Similarity 57.1%; Pred. No. 9.93e-01;

Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 72 HTGLTRRGCGT 85
:::|||||::|
Qy 1 QSGTMRTRHSTGCT 14

RESULT 8
ENTRY 151592 #type complete
TITLE p59(Xfyn) - Xiphophorus helleri
#formal_name Xiphophorus helleri
ORGANISM
DATE 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 18-Jun-1999

ACCESSIONS 151592
REFERENCE 151592
#authors Hannig, G.; Ottillie, S.; Scharf, M.
#journal Oncogene (1991) 6:361-369
#title Conservation of structure and expression of the c-yes and fyn genes in lower vertebrates.
#cross-references MUID:9187435
#accession 151592
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-537 #label HAN
#cross-references EMBL:X54971; NID:g64481; PIDN:CAA38715.1; PID:g64482

GENETICS
#gene Xfyn
#superfamily protein-tyrosine kinase src: protein kinase
CLASSIFICATION
#homology: SH2 homology; SH3 homology

FEATURE
89-138 #domain SH3 homology #label SH3\
149-246 #domain SH2 homology #label SH2\
269-527 #domain protein kinase homology #label KIN
SUMMARY
#length 537 #molecular-weight 60447 #checksum 621
Query Match 55.6%; Score 55; DB 2; Length 537;
Best Local Similarity 57.1%; Pred. No. 9.93e-01;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 72 HTGLTRRGCGT 85
:::|||||::|
Qy 1 QSGTMRTRHSTGCT 14

RESULT 9
ENTRY A43806 #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) fyn - African clawed frog
ORGANISM
DATE 30-Jan-1993 #sequence_revision 30-Jan-1993 #text_change 18-Jun-1999

ACCESSIONS A43806
REFERENCE A43806
#authors Steele, R.E.; Deng, J.C.; Ghosh, C.R.; Fero, J.B.
#journal Oncogene (1990) 5:369-376
#title Structure and expression of fyn genes in Xenopus laevis.
#cross-references MUID:90191723
#accession A43806
#status preliminary
#molecule_type mRNA
#residues 1-537 #label STE
#cross-references GB:X52188; GB:X52189; NID:g64701; PIDN:CAA36435.1; PID:g64702

CLASSIFICATION
#superfamily protein-tyrosine kinase src: protein kinase
#homology: SH2 homology; SH3 homology
KEYWORDS
ATP; autophosphorylation; phosphoprotein; phosphotransferase; tyrosine-specific protein kinase

FEATURE
89-138 #domain SH3 homology #label SH3\
149-246 #domain SH2 homology #label SH2\
269-527 #domain protein kinase homology #label KIN\
277-285 #region protein kinase ATP-binding motif
SUMMARY
#length 537 #molecular-weight 60846 #checksum 2563

Query Match 55.6%; Score 55; DB 2; Length 537;
Best Local Similarity 57.1%; Pred. No. 9.93e-01;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 72 HTGTRTGTGTGT 85
OY 1 OSCTMRHSTGCT 14

RESULT 10
ENTRY 70369 #type complete
TITLE gene m3 protein - Lactococcus phage b1170
ORGANISM #formal_name Lactococcus phage b1170
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999

ACCESSIONS 70369
REFERENCE 214903
#authors Cruz-Le Cocq, A.M.; Cesselin, B.; Commissaire, J.; Anba, J.; Kyriakidis, S.; Chopin, M.C.
#submission Submitted to the EMBL Data Library, June 1997
#description Sequence and organization of the lactococcal isometric b1170 phage genome.
#accession T03369
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-160 #label CRU
GENETICS #cross-references EMBL:AF009630; NID:g3282260; PID:g3282321

SUMMARY #gene m3
#length 160 #molecular-weight 18072 #checksum 3777

Query Match 53.5%; Score 53; DB 2; Length 160;
Best Local Similarity 60.0%; Pred. No. 2.63e+00;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 90 GVMRNRHNG 99
OY 3 GTMRTRHSTG 12

RESULT 11
ENTRY 713827 #type complete
TITLE kinesin-73 - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

ACCESSIONS 713827
REFERENCE 21784
#authors Li, H.P.; Liu, Z.M.; Nirenberg, M.
#journal Proc. Natl. Acad. Sci. U.S.A. (1997) 94:1086-1091
#title kinesin-73 in the nervous system of Drosophila embryos.
#accession 713827
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-1921 #label LTH
#cross-references EMBL:U01788; NID:g1906595; PID:g1906596; PDB:1AAB50404.1

SUMMARY #length 1921 #molecular-weight 215047 #checksum 9262

Query Match 52.5%; Score 52; DB 2; Length 1921;
Best Local Similarity 58.3%; Pred. No. 4.24e+00;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1431 GGVRSRHSFGG 1442
OY 2 SCTMRTRHSTG 13

RESULT 12
ENTRY JN0578 #type complete
TITLE p-Aminobenzoic acid synthetase (EC 4.1.3.-) B - Streptomyces lividans

ORGANISM #formal_name Streptomyces lividans
DATE 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 22-Jun-1999

ACCESSIONS JN0578
REFERENCE JN0577
#authors Arlin, F.F.; Vining, L.C.
#journal Gene (1993) 126:129-133
#title Organization of the genes encoding p-aminobenzoic acid synthetase from Streptomyces lividans 1326.
#cross-references MUID:93231526
#accession JN0578
#molecule_type DNA
#residues 1-475 #label ARH
#cross-references GB:M64059; NID:g153393; PID:AAA26798.1; PID:g153394
COMMENT This enzyme participates in p-aminobenzoic acid biosynthesis.
GENETICS
#gene PabB
CLASSIFICATION #superfamily anthranilate synthase component I
KEYWORDS carbon-carbon lyase; oxo-acid-lyase; p-aminobenzoate biosynthesis

SUMMARY #length 475 #molecular-weight 52042 #checksum 6682

Query Match 51.5%; Score 51; DB 2; Length 475;
Best Local Similarity 42.9%; Pred. No. 6.79e+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 281 ACTLGRPRAGCD 294
OY 2 SCTMRTRHSTGCTN 15

RESULT 13
ENTRY S04605 #type complete
TITLE glycyl-in G3 - soybean
ORGANISM #formal_name Glycine max #common_name soybean
DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 20-Aug-1999

ACCESSIONS S04605
REFERENCE S04605
#authors Cho, T.J.; Nielsen, N.C.
#journal Nucleic Acids Res. (1989) 17:4388
#title The glycyl-in G(3) gene from soybean.
#cross-references MUID:89296500
#accession S04605
#molecule_type DNA
#residues 1-481 #label CHO
#cross-references EMBL:X15123; NID:g18638; PID:CAA33217.1; PID:g18639
#experimental_source variety Date

GENETICS
#gene GY3
#introns 96/1: 177/3: 352/3
CLASSIFICATION #superfamily glycyl-in
SUMMARY #length 481 #molecular-weight 54241 #checksum 2269

Query Match 51.5%; Score 51; DB 2; Length 481;
Best Local Similarity 58.3%; Pred. No. 6.79e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 304 TMRNRHNGTIS 315
OY 4 TMRTRHSTGCTN 15

RESULT 14
ENTRY S11003 #type complete
TITLE glycyl-in G3 precursor - soybean
ORGANISM #formal_name Glycine max #common_name soybean
DATE 21-Nov-1993 #sequence_revision 19-Jan-1996 #text_change 03-May-1996

ACCESSIONS S11003
REFERENCE S10851
#authors Nielsen, N.C.; Dickinson, C.D.; Cho, T.J.; Thanh, V.H.; Scallan, B.J.; Fischer, R.L.; Sims, T.L.; Drews, G.N.;

Goldberg, R.B.
#journal Plant Cell (1989) 1:313-328
#title Characterization of the glycylcin gene family in soybean.
#cross-references MUID:92393391
#accession S11003
#molecule_type DNA
#residues 1-484 #label NIE
##experimental_source variety Dare

GENETICS
#gene Gy3
CLASSIFICATION #superfamily glycylcin
KEYWORDS storage protein
FEATURE
1-19
20-484
SUMMARY
#domain signal sequence #status predicted #label SIG\
#product glycylcin G3 #status predicted #label MAT\
#length 484 #molecular_weight 54359 #checksum 8616

Query Match 51.5%; Score 51; DB 2; Length 484;
Best Local Similarity 58.3%; Pred. No. 6.79e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 304 TMRLRHNGOTS 315
||| ||: ||:
QY 4 TMRTRHSTGCTN 15

RESULT 15
ENTRY FMSYC2 #type complete
TITLE glycylcin chain A1abx precursor - soybean
ALTERNATE_NAMES 115 globulin; glycylcin A1ab1b
ORGANISM #formal_name glycine max #common_name soybean
DATE 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jul-1999
A23497: S10502

ACCESSIONS
REFERENCE A23497
#authors Negoro, T.; Momma, T.; Fukazawa, C.
#journal Nucleic Acids Res. (1985) 13:6719-6731
#title A cDNA clone encoding a glycylcin A-1a subunit precursor of soybean.
#cross-references MUID:86041867
#accession A23497
#molecule_type mRNA
#residues 1-495 #label NEG
##cross-references GB:X02985; NID:q18614; PIDN:CAA26723.1; PID:q18615
##experimental_source cv. Bomblori
##note the authors translated the codon AAC for residue 449 as lys
##note because of current nomenclature ambiguities, the authors choose to designate the basic chain as Bx

REFERENCE S10502
#authors Kitamura, Y.; Arahira, M.; Itoh, Y.; Fukazawa, C.
#journal Nucleic Acids Res. (1990) 18:4245
#title The complete nucleotide sequence of soybean glycylcin A2B1a gene spanning to another glycylcin gene A1ab1b.
#cross-references MUID:90332420
#accession S10502
#status preliminary: translation not shown
#molecule_type DNA
#residues 481-495 #label KIT
##cross-references EMBL:X53404; NID:q18522; PIDN:CAA37479.1; PID:q18523
COMMENT The source of this protein was cotyledon tissue taken 38 days after flowering.
COMMENT By homology with the A2B1a component, residues 307-310, and 491-495 are removed from the precursor during posttranslational processing.

CLASSIFICATION #superfamily glycylcin
KEYWORDS seed; storage protein
FEATURE
1-19 #domain signal sequence #status predicted #label SIG\
20-306 #product glycylcin chain A1a #status predicted #label GLA\
311-490 #product glycylcin chain Bx #status predicted #label GLB\
107-317 #disulfide_bonds #status predicted

SUMMARY #length 495 #molecular_weight 55506 #checksum 8518
Query Match 51.5%; Score 51; DB 1; Length 495;
Best Local Similarity 58.3%; Pred. No. 6.79e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 318 TMRLRHNGOTS 329
||| ||: ||:
QY 4 TMRTRHSTGCTN 15

Search completed: Sat May 13 07:36:42 2000
Job time : 8 secs.

 WISE (TM)

Release 3.1A John F. Collins, Bioinformatics Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 07:29:58 2000; Maspar time 89.82 Seconds

Tabular output not generated. 5.086 Million cell updates/sec

Title: >US-08-991-628-7

Description: (1-15) from US08991628.pep

Sequence: 1 OSGTMRTRHSTGCTN 15

Scoring table: PAM 150
Gap 15

Searched: 83857 segs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 24.519; Variance 27.183; scale 0.902

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|------------|------------------------|-----------|
| 1 | 99 | 100.0 | 999 | 1 | DSG3_HUMAN | DESMOGLEIN 3 PRECURSOR | 3.88e-13 |
| 2 | 57 | 57.6 | 1597 | 1 | SOL_DROME | SMALL OPTIC LOBES PROT | 1.21e-01 |
| 3 | 56 | 56.6 | 478 | 1 | RPL1_EUPOC | DNA-DIRECTED RNA POLYM | 2.08e-01 |
| 4 | 55 | 55.6 | 533 | 1 | FYN_MOUSE | PROTO-ONCOGENE TYROSIN | 3.55e-01 |
| 5 | 55 | 55.6 | 533 | 1 | FYN_MOUSE | PROTO-ONCOGENE TYROSIN | 3.55e-01 |
| 6 | 55 | 55.6 | 533 | 1 | FYN_MOUSE | PROTO-ONCOGENE TYROSIN | 3.55e-01 |
| 7 | 55 | 55.6 | 533 | 1 | FYN_MOUSE | PROTO-ONCOGENE TYROSIN | 3.55e-01 |
| 8 | 55 | 55.6 | 533 | 1 | FYN_MOUSE | PROTO-ONCOGENE TYROSIN | 3.55e-01 |
| 9 | 52 | 52.5 | 543 | 1 | TCPH_HUMAN | T-COMPLEX PROTEIN 1, E | 1.69e+00 |
| 10 | 51 | 51.5 | 475 | 1 | PAB3_STRLI | PARA-AMINOENZOAMT SYN | 2.80e+00 |
| 11 | 51 | 51.5 | 481 | 1 | GLC1_SOYBN | GLYCININ G3 PRECURSOR | 2.80e+00 |
| 12 | 51 | 51.5 | 495 | 1 | GLC1_SOYBN | GLYCININ G1 PRECURSOR | 2.80e+00 |
| 13 | 50 | 50.5 | 120 | 1 | MASD_SCHUA | IMMUNOGENIC MIRACIDIAL | 4.61e+00 |
| 14 | 50 | 50.5 | 286 | 1 | AAC8_STRFR | AMINOGLYCOSIDE N3'-ACE | 4.61e+00 |
| 15 | 50 | 50.5 | 572 | 1 | HEMA_P13B | HEMAGGLUTININ-NEURAMIN | 4.61e+00 |
| 16 | 50 | 50.5 | 2318 | 1 | NTC3_MOUSE | NEUROGENIC LOCUS NOTCH | 4.61e+00 |
| 17 | 49 | 49.5 | 230 | 1 | RL4_MYCLE | 50S RIBOSOMAL PROTEIN | 7.54e+00 |
| 18 | 48 | 48.5 | 172 | 1 | SCVD_MAGGR | SCYALONE DEHYDRATASE | 1.22e+01 |
| 19 | 48 | 48.5 | 291 | 1 | T2EB_HUMAN | TRANSCRIPTION INITIATI | 1.22e+01 |
| 20 | 48 | 48.5 | 355 | 1 | COAT_AEDER | COAT PROTEIN VP1 [CONT | 1.22e+01 |
| 21 | 48 | 48.5 | 462 | 1 | BIND_LYTV4 | BINDIN PRECURSOR. | 1.22e+01 |
| 22 | 48 | 48.5 | 471 | 1 | NRAM_IAPFW | NEURAMINIDASE (EC 3.2. | 1.22e+01 |
| 23 | 48 | 48.5 | 510 | 1 | KR2_VAVD | PROBABLE SERINE/THREON | 1.22e+01 |

| | | | | | | | |
|----|----|------|------|---|------------|------------------------|----------|
| 24 | 48 | 48.5 | 510 | 1 | CATB_PSESY | CATALASE PRECURSOR (EC | 1.22e+01 |
| 25 | 48 | 48.5 | 729 | 1 | NARB_SYNP7 | NITRATE REDUCTASE (EC | 1.22e+01 |
| 26 | 48 | 48.5 | 823 | 1 | UBPG_HUMAN | UBIQUITIN CARBOXYL-TER | 1.22e+01 |
| 27 | 48 | 48.5 | 927 | 1 | CC15_SCHPO | CELL DIVISION CONTROL | 1.22e+01 |
| 28 | 48 | 48.5 | 1053 | 1 | CAPP_ANANI | PHYTOCHROME A. | 1.22e+01 |
| 29 | 48 | 48.5 | 1125 | 1 | PHYA_POPTM | PHYTOCHROME A. | 1.22e+01 |
| 30 | 48 | 48.5 | 1130 | 1 | C2TA_HUMAN | MHC CLASS II TRANSACTI | 1.22e+01 |
| 31 | 48 | 48.5 | 1201 | 1 | COPA_YEAST | CONTOMER ALPHA SUBUNIT | 1.22e+01 |
| 32 | 48 | 48.5 | 1591 | 1 | TIAM_MOUSE | T-LYMPHOMA INVASION AN | 1.22e+01 |
| 33 | 48 | 48.5 | 1591 | 1 | TIAM_MOUSE | T-LYMPHOMA INVASION AN | 1.22e+01 |
| 34 | 48 | 48.5 | 220 | 1 | RS8_ORYSA | 40S RIBOSOMAL PROTEIN | 1.97e+01 |
| 35 | 47 | 47.5 | 221 | 1 | RS8_MOUSE | 40S RIBOSOMAL PROTEIN | 1.97e+01 |
| 36 | 47 | 47.5 | 370 | 1 | MAF2_MOUSE | TRANSCRIPTION FACTOR M | 1.97e+01 |
| 37 | 47 | 47.5 | 416 | 1 | FI1A_HUMAN | FLIAGRIN PRECURSOR | 1.97e+01 |
| 38 | 47 | 47.5 | 469 | 1 | NRAM_TAHCO | NEURAMINIDASE (EC 3.2. | 1.97e+01 |
| 39 | 47 | 47.5 | 469 | 1 | NRAM_TAHCO | NEURAMINIDASE (EC 3.2. | 1.97e+01 |
| 40 | 47 | 47.5 | 470 | 1 | NRAM_TAHCO | NEURAMINIDASE (EC 3.2. | 1.97e+01 |
| 41 | 47 | 47.5 | 512 | 1 | AMY_BAGLI | ALPHA-AMYLASE PRECURSO | 1.97e+01 |
| 42 | 47 | 47.5 | 544 | 1 | TCPH_MOUSE | T-COMPLEX PROTEIN 1, E | 1.97e+01 |
| 43 | 47 | 47.5 | 606 | 1 | POUG_YEYF8 | DNA TERMINAL PROTEIN (| 1.97e+01 |
| 44 | 47 | 47.5 | 1163 | 1 | FAB1_YEAST | GENOME POLYPROTEIN (CO | 1.97e+01 |
| 45 | 47 | 47.5 | 2278 | 1 | FAB1_YEAST | PROBABLE PHOSPHATIDYL | 1.97e+01 |

ALIGNMENTS

RESULT 1
ID DSG3_HUMAN STANDARD: PRT: 999 AA.

AC P32926;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;

OC Eukaryota; Primates; Catarrhini; Hominoidea; Homo.

RP [1]

RX MEDLINE; 92069753.

RA Amagai M., Klaus-Kovtun V., Stanley J.R.;

RT "Autoantibodies against a novel epithelial cadherin in pemphigus

RT vulgaris, a disease of cell adhesion.";

RL Cell 67:869-877(1991).

CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.

CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE

CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND

CC CARCINOMAS.

CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS

CC (POTENTIAL).

CC -1- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN

CC DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE

CC LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES

CC AGAINST DSG3.

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation at

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL; M76482; AAA60230.1; -.

DR PIR; A41088; IYHUG3.

DR HSSP; P09803; 1EDH.

DR MKI; 169615; -.

DR PROSITE; PS00232; CADHERIN; 3.

DR PIR; P00028; cadherin; 4.

Cell adhesion; signal; Transmembrane; Cytoskeleton; Glycoprotein;

THIS PAGE BLANK (USPTO)